UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

FACULDADE DE LETRAS

FACULDADE DE MEDICINA

FACULDADE DE PSICOLOGIA

Aging effects in speech statistical learning - a behavioral and computational study

Bruno Rafael Gil Penha

Dissertação de Mestrado

MESTRADO EM CIÊNCIA COGNITIVA

2014

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

FACULDADE DE LETRAS

FACULDADE DE MEDICINA

FACULDADE DE PSICOLOGIA

Aging effects in speech statistical learning - a behavioral and computational study

(Efeitos do envelhecimento na aprendizagem estatística do discurso - um estudo comportamental e computacional)

Bruno Rafael Gil Penha

Dissertação orientada pela Profª. Doutora Tânia Fernandes
e pelo Prof. Doutor Luís Correia

Dissertação de Mestrado

MESTRADO EM CIÊNCIA COGNITIVA

2014

O conteúdo desta dissertação reflete as perspetivas, o trabalho e as interpretações do autor no momento da sua entrega. Esta dissertação pode conter incorreções, tanto conceptuais como metodológicas, que podem ter sido identificadas em momento posterior ao da sua entrega. Por conseguinte, qualquer utilização dos seus conteúdos deve ser exercida com cautela. Ao entregar esta dissertação, o autor declara que a mesma é resultante do seu próprio trabalho, contém contributos originais e são reconhecidas todas as fontes utilizadas, encontrando-se tais fontes devidamente citadas no corpo do texto e identificadas na secção de referências. O autor declara, ainda, que não divulga na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor ou de propriedade industrial.

To my parents

who grew old too fast because of me

Acknoledgements

I would like to thank my supervisers Tânia and Luís, that had more patience for me than what I thought was humanely possible. I am also in great debt to Rob for his advises and help with the simulations.

A very big thanks to all the participants of this study: students from the Faculdade de Psicologia da Universidade de Lisboa and students from the Universidade Internacional para a Terceira Idade (UITI) and Universidade Sénior na Universidade (UNIDAC). Also a special thanks to Mª da Ascenção Rebelo and Mônica Vieira, the directors of these marvelous institutions, for the kindness in letting me promote the study among the students and use the institutions' facilities for the study.

My thanks go also to the Center for Research in Psychology at the University of Lisboa (CIPUL) for helping me find participants and allowing me to test them in the faculty's labs and to the Neurocognition and Language Research Group, Center for Psychology at University of Porto and to its coordinator Prof. São Luís Castro for providing me some of the data.

Without the companionship that emerged between the members of the Cognitive Sciences Master's this journey would have not been so enjoyable. It was truly a pleasure to meet and exchange ideas with such talented minds.

And above all I want to thank my parents that in the course of almost a quarter of century have supported me in the pursuit of my dreams.

Resumo

A segmentação da fala é um empreendimento dificil devido à falta ou inconsistencia de marcadores directos de fronteiras entre palavras, tal como pausas. No entanto, o discurso falado é altamente estruturado, originando co-ocorrencias entre estímulos, em relação aos quais os sujeitos são sensiveis e capazes de adquirir através de um processo conhecido por aprendizagem estatística. A aprendizagem estatística pode ocorrer através de diversos tipos de computações entre estímulos, sendo um destes a probabilidade transicional (TP, do inglês *transitional probability*) de sílabas sequenciais. Estudos sobre a segmentação da fala recorrendo a TPs entre sílabas em crianças e jovens adultos demonstram que estas sozinhas são suficientes para segmentar o discurso em línguas artificiais. No entanto, estudos do género ainda não foram realizados numa população mais velha.

Neste trabalho comparamos o desempenho de jovens adultos e adultos mais velhos num paradigma de aprendizagem de língua artificial. Foi usada uma lingua artificial cuja única pista para a segmentação consistia em TPs: sílabas entre palavras apresentam uma TP pequena enquanto sílabas pertencentes a uma palavra têm uma TP elevada, simulando as línguas naturais. Após a exposição à lingua os participantes fizeram um teste de escolha forçada entre palavras e parte-palavras (segmento da língua abarcando pedaços de duas palavras), onde foi pedido aos participantes que identificassem as palavras pertencentes à lingua. Tanto as palavras como as parte-palavras foram divididas nas categorias alta-PT e baixa-PT, permitindo uma análise mais detalhada.

A aprendizagem estatística está dependente do foco da atenção para um desempenho eficaz. Devido ao facto dos jovens adultos terem um mau desempenho na escolha de palavras de baixa-PT quando fazem tarefas de divisão de atenção, esperamos resultados semelhantes para os participantes mais velhos nesta tarefa devido as declínios neuronais característicos do envelhecimento em zonas dos cérebro importantes para a aprendizagem estatística e a atenção.

Dois modelos de segmentação da fala, PARSER e Redes Recurrentes Simples, foram testados de mode a avaliar quão próximos estão do desempenho humano. O PARSER baseia-se numa segmentaçao por *chunks* enquanto que as Redes Recurrentes Simples se baseiam nas TPs.

Os nossos resultados comportamentais demonstram que ambos os grupos têm um desempenho semelhante na identificação de palavras de alta-TP mas os participantes mais velhos têm um desempenho significativamente pior na identificação de palavras de baixa-TP enquanto os participantes jovens surpreendentemente têm um melhor desempenho em comparação com as palavras alta-PT. Testes auxiliares permitiram teorizar a razão por detrás destes resultados, mas estudos futuros de neuroimagem podem dar respostas mais concretas a este problema.

Os resultados computacionais demonstram que o PARSER supera em muito os participantes humanos, enquanto o desempenho da SRN é muito superior nas palavras de alta-TP mas é semelhante nas palavras de baixa-TP. Apesar da simplicidade dos modelos usados, os resultados parecem indicar que a computação das TPs tem uma maior credibilidade na segmentação de fala do que os processos aletórios de escolha de *chunks* e mecanismos de memória associados.

Palavras Chave:

Segmentação do Discurso; Aprendizagem Estatística; Envelhecimento; Aprendizagem de Lingua Artificial; Modelação Computacional

Abstract

Speech segmentation is a difficult enterprise due to the lack or inconsistency of direct marking of word boundaries in speech, such as pauses. However, speech stream is highly structured, leading to co-ocurrences between stimuli, which learners are sensitive to and can acquire through a process known as statistical learning (SL). Cues to speech segmentation can be prosodic, acoustic-phonetic or distributional, this last class consisting of several computations, such as transitional probabilities (TP) between syllables. Studies on syllables TP have shown that they alone are sufficient to segment speech. Although there have been some studies about the importance of syllables TPs in speech segmentation in infants and young adults, this kind of research is lacking in the elderly.

In this work we compare young and older adult performance in an artificial language learning paradigm. An artificial language that excludes other than distributional cues was used, enabling the exploration of this kind of cues importance for speech segmentation. After the participants were exposed to the language a forced-choice test was enforced between words and part-words where participants were asked to identify the word belonging to the language. Both words and part-words were divided in high-TP and low-TP categories, allowing for a fine-grained analysis.

SL is dependent on attentional focus for an effective performance. Because young adults perform poorly at choosing low-TP words while doing a divided attention task, we expect similar results for older participants in this task due to brain areas' important for SL and attentional declines characteristic of aging.

Two computational models of word segmentation, PARSER and Simple Recurrent Network (SRN), were tested to see how close they are from human performance. Each represents a different approache to speech segmentation: PARSER is chunk-based while SRN is TP-based

Our behavioral results show that both groups perform similarly in identifying high-TP words but older participants performed significantly worse (at chance level) in identifying low-TP words whether young participants surprisingly performed better compared to high-TP. Because this study did not include a neuroimaging component, we can only theorize the reason behind these results. Further neuroimaging studies could give a more concrete answer to this problem.

Our computational results show that PARSER outperforms humans by far, while SRN performance is more alike humans'. Despite the simplicity of the models here used, results seem to indicate that TP computations have a greater credibility in speech segmentation than random choosing of chunks and further memory mechanisms.

Key Words:

Speech Segmentation; Statistical Learning; Aging; Artificial Language Learning; Computational Modeling

Index

Figure Index

Table Index

Appendix Index

Acronyms

Artificial Grammar Learning	AGL
Artificial Language	AL
Artificial Language Learning	ALL
Reaction Time	RT
Serial Reaction Time	SRT
Simple Recurrent Network	SRN
Statistical Learning	SL
Transitional Probability	TP
Two-alternative Forced Choice	2AFC

Behavioral Experiment

1. Introduction

Speech word segmentation is commonly believed to be achieved by infants and young adults alike through statistical learning (Saffran et al. 1996a; Saffran et al. 1996b; Aslin et al. 1998). Statistical learning (SL) refers to the process of discovering and identifying features of the input that predict other features, grouping those likely to co-occur. Learners are sensitive to different features of the input and are able to compute different statistics. Transitional probabilities (TP) have been the most studied statistic in the SL literature (Thiessen 2013). The TP between two items X and Y is the number of times XY occurs divided by the number of occurrences of X. In natural speech adjacent within-word syllables have generally higher TP than adjacent syllables that cross a word boundary (e.g., Perruchet & Peereman, 2004). The simulation of this characteristic in artificial languages led to the conclusion that both infants and adults exploit TPs to group items in the input that are likely to co-occur (Aslin et al. 1998), successfully segmenting this way words in speech (Saffran et al. 1996a; Saffran et al. 1996b).

There is however one important period of life that is underrepresented in SL literature: that of late adulthood. Since the older population is increasingly taking a bigger slice of the world demographics, it should not be underrepresented in SL research. Further investigation on this phenomenon would not only have scientific value but could also help to develop practical application on healthy aging.

Natural cognitive aging results in a decline in structure and function of brain regions relevant for SL (e.g. Raz et al, 2005; Salthouse, 2010), including those responsible for attentional allocation (Zanto & Gazzaley, 2014) and working memory (Braver and West, 2008; Borella et al., 2008), which would, in theory, lead to a lower performance of older adults compared to younger adults. It is then relevant to test if older adults' performance in a speech SL task reflects a decline compared to young adults).

Despite the evidence supporting SL as an important contributor to speech segmentation, some authors remain skeptical about its importance for word segmentation, attributing the main role to other processes, such as chunking (e.g. Perruchet & Vinter, 1998). A chunk is a set of perceptual primitives, such as syllables, with strong associations between them and weaker associations with elements of other chunks. According to Perruchet & Vinter (1998), word segmentation results from the chunking of attentionally relevant perceptual primitives into increasingly bigger chunks, eventually reaching the size of a word. Every time a chunk appears in the input its weight in memory would be strengthened. All these chunks would then compete for memory resources and, over time, the most salient of these would be words. To prove their point these authors developed Parser, a computational model that has been having reasonable success in fitting human performance in word segmentation tasks (Perruchet & Vincent 2014). The Simple Recurrent Network (SRN) is the computational model commonly used to implement the TP-approach. Computational modeling can help to assess the plausibility of each of these approaches. Still, there are very few studies comparing the performance of each model with the data from the same behavioral study. The possibility of doing so with the data from this behavioral study, including data from older adults, never used before, would enable a deeper understanding of the processes behind word segmentation.

Since this thesis is the result of a Cognitive Science masters, we belief that it should be a crossing road between several disciplines. The present study is then divided in two experiments: a behavioral experiment in the field of psycholinguistics that provided the data for a consequent neurocomputational experiment.

In the behavioral experiment we explored the effects of cognitive aging in the performance of a speech statistical learning task. In chapter 1 we develop the theoretical background involving the segmentation problem, statistical learning and cognitive aging. In chapter 2 we describe in detail the methods employed in this experiment. In chapter 3 we analyze the results and briefly discuss them.

In the computational experiment we explore how well does the performance of a SRN model developed by us and PARSER fit to the human behavioral data. In chapter 4 we introduce the reader to computational modeling, we describe in detail each model, their parameters and how we reached their value in order to account for the young and older adults' performance. In chapter 5 we describe the methods and discuss the results. In chapter 6 we do a general discussion of the results of both experiments.

In chapter 7 we present the conclusions of this study, its limitations and wonder about the next direction to take in the exploration of speech statistical learning in older adults.

1.1 The Segmentation Problem

The first challenge that infants and second-language learners have to overcome when acquiring a language is the identification of the units that compose the speech signal. Known as the segmentation problem, discovering the location of word boundaries in speech is not straightforward since they lack direct marking (i.e. the equivalent of white spaces in written text). However, different sources of information can be explored by the learners in order to accomplish speech segmentation. Among these are prosodic (i.e. the intonation of speech), stress and phonotactic (i.e. the constraints on the occurrence of sounds within words and sentence of a given language) patterns (Mattys et al., 2005). Humans seem to be able to unconsciously learn these patterns in speech through the computation of different statistics between the units of speech (phonemes or syllables), such as frequency count or conditional probabilities (Romberg & Saffran, 2010). This process was termed as *statistical learning*[1] *(SL).*

[1] Not to be confused with Statistical Learning Theory, a theoretical branch of machine learning that deals with the problem of induction and causal inference (i.e. extraction of rules from a small set of event in order to predict the occurrence of the following events).

Alternatively some authors use the term *sequential learning* if the input units of the task in question occur sequentially. Since it was later found that humans can do the same kind of learn in another modalities, it is now common to use the term speech statistical learning when this type of learning has to do with speech.

1.2 Speech statistical learning as the solution

Speech SL can be divided into three phases: firstly, the learner has to keep track of different auditory stimuli and compute the interactions between them; secondly, auditory stimuli and computation have to be encoded; finally, these computations have to be retrieved in order to be used.

Typical neurons fire specifically to particular subsets of phonemes (Chan et al., 2013). The co-occurrence of two phonemes would imply a co-occurrence of neurons firing. According to associative or Hebbian learning, there is an increase in the strength of the connections between adjacent neurons that have a causal relation of firing, being this connections responsible for holding the information (Hebb 1949). The first and second phases of SL might be solely explained by this mechanism. Instead of passive learning, current studies are more supportive of the theory of the brain as a sophisticated hypothesis tester that seeks to minimize the error between its predictions of what the sensory input would be and the actual incoming input (Gregory, 1980; Friston, 2005). In accordance with this theory, the learner is assumed to represent a set of hypotheses (the hypotheses space) about word boundaries and then to accept or reject it along the learning process. There are an unlimited number of hypotheses for the hypotheses space of word boundaries, which indicates that some constraints must be in place to narrow it. The abilities to select relevant information (attention) and to retain information in an accessible state (working memory) could be a solution to the combinatorial explosion.

The mechanisms behind the last phase are also not clear. The ease with which an item is retrieved from memory is dictated, at least in part, by other items similar to the target (e.g.

Marsh et al 2008; Anderson 2003). This construct is called *interference,* and has been considered one of the causes of forgetting by classical structuralism that see forgetting as a side-effect of structural changes that result from the storage of new items in memory (e.g. Anderson, 1983). Once again selective attention is important to inhibit irrelevant information, optimizing the efficiency of working memory and keeping interference at a minimum.

The initial studies of speech SL began by exploring how the information contained in the transitional probabilities (TP) between syllables in artificial speech enabled speech segmentation (Saffran et al. 1996a; 1996b). A TP between two syllables is the conditional probability of a syllable A predicting the syllable B such that TP(A/B) = P(AB)/P(A), where P(AB) = frequency of B following A, and P(A) = frequency of A. In natural languages adjacent syllables within a same word have generally higher TPs than adjacent syllables that cross a word boundary (Perruchet & Peereman, 2004). Consider, as an example, the English utterance “pretty baby”: while the probability of “ty” following “pre” is 80% in infant-directed English speech, the probability of “ba” following “ty” is a mere 0.03% in the same context (Saffran, 2003). In this case “ty” strongly predicts that “pre” will occur, so the TP between these two syllables is high. Saffran et al. (1996a; 1996b) demonstrated, through the application of an artificial language learning paradigm where the only cues available for the segmentation of the artificial language were phonotactic, that both infants and adults were able to segment the continuous speech signal into word-like units.

This paradigm was subsequently modified to explore other aspects of SL, leading to the conclusion that SL contributes to the acquisition of other levels of language, such as lexicon development (Yu and Ballard, 2007), and to the acquisition of input regularities in other modalities, such as touch and vision (Conway & Christian, 2005; Fiser & Aslin 2002; Kirkham et al., 2002). SL has even been demonstrated in nonhuman primates (Heimbauer et al., 2010) and rodents (Toro & Trobalon, 2005) suggesting a phylogenetic ancient root. Such a widespread use of SL in perception led to the suggestion that it is a domain-general phenomenon used in different modalities (e.g., Kirkham et al., 2002; Bapi et al. 2005). Others suggested that SL might have multiple coexisting subsystems with similar computational principles, each better suited for a specific task (e.g. Goschke, 1998; Seger, 1997; Emberson et al., 2011). Emberson et al (2011) presented auditory and visual statistically equivalent cues to young adult participants at different rates and concluded that while at faster presentation rates auditory SL was superior to visual SL, at slower presentation the opposite occurred. It is also possible that both domain-general and modality-specific SL mechanisms might coexist (Keele et al., 2003; Conway and Christiansen, 2005; Conway and Pisoni, 2008; Turk-Browne et al., 2009; Shafto et al., 2012). A theory by Keele and colleges (2003) posits that the human brain supports two broad systems of sequence learning: a multidimensional and a unidimensional systems with different neural pathways and attentional requirements.

Whatever the case might be, one should be cautious if it is to extrapolate results between studies of SL in different conditions.

1.3 The artificial language learning paradigm

The complexity of natural languages makes it difficult to study SL in the lab. Studies of speech SL resort to artificial languages learning (ALL) paradigms to isolate a particular cue to segmentation and control the statistical value of the interactions between the units of speech. The ALL is classically divided into a listening (or familiarization) phase and a test phase. On adults, the paradigm was first applied by Saffran et al. (1996b), remaining unchanged until today. It will be described here as a typical instantiation of the paradigm. In the familiarization phase learners were exposed to an AL stream of continuous, nonsensical speech during 21min.

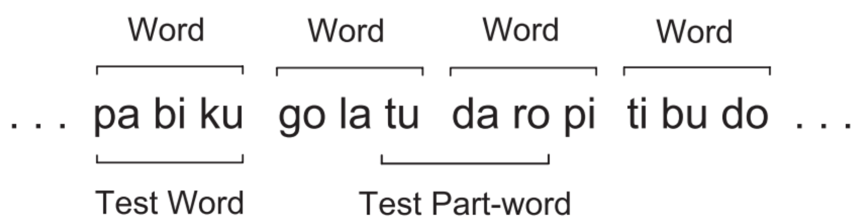


Figure 1 - Example of a test word and a test part-word from Saffran et al. (1996b) artificial language. Adapted from Aslin & Newport (2012).

This AL was constructed by concatenating several instances of 6 trisyllabic words (see Figure 1). The TPs of adjacent syllables within a word were 1.0, while TPs of adjacent syllables spanning a word boundary were 0.33. In the second phase, the test phase, participants listened to two stimuli, one word and one part-word, and were asked to choose the stimulus they believe to be a word of the language. This procedure is known as a two-alternative forced-choice (2AFC) test. A part-word is a trisyllabic sequence that spanned word boundaries (see Figure 1). Six part-words were chosen from all possible part-words so that there is an equal number of words and part-words to be matched against each other in the 2AFC. Participant performance, the average of all the 6x6 comparisons between words and part-words, was then compared to chance level (i.e. 50% of correct trials), and if it was statistically superior to this value then it was assumed that a (somewhat, depending on the value) successful segmentation occurred.

Recent studies using this paradigm have tried to approximate the experiments in speech SL to real-world language learning, for example by increasing the complexity of the ALs to better match that of natural languages. This can be accomplished by, for example, increasing the variation of TPs between syllables from the traditional values of 1 and .33. That is the case of the AL used by Fernandes et al. (2007) that displays a wider range of TPs between syllables. The words of the AL were divided into two classes, namely low-TP-words and high-TP-words, depending on whether the average TPs between the syllables of each word was around .5 or higher than .75, respectively. In the 2AFC test, high-TP-words were correctly picked more often than low-TP-words by young adults (Fernandes et al. 2007). This study allowed for a finer inspection of the TPs values importance in the process of speech segmentation. Another important approximation to real-world language learning that has been missing is the investigation of SL in more age groups than just infants and young adults. Investigation on late adulthood, for instance, has been missing. Since the older population is increasingly taking a bigger slice of the world demographics, it should not be underrepresented in SL research. Further investigation on this phenomenon would not only have scientific value but could also help to develop practical application on healthy aging.

1.4 Speech statistical learning in late adulthood

Several studies established that speech SL remains a viable mechanism of speech segmentation in infancy and young adulthood (Saffran et al, 1996a; 1996b). Speech SL studies in late adulthood, however, are missing. A rare example is the study of Schapiro et al. (2014) where the authors tested the performance of 28 healthy participants with a mean age of 62.9 years in four SL conditions: shapes, syllables, scenes, or tones. Each participant was assigned to one of the tasks. Each task had 12 items organized into 4 triplets, randomly concatenated into a total of 288 items, except for the syllables task that had 1152 items, all with a duration of 4.8 min. In the test phase participants completed 32 2AFC trials where one of the triplets were paired with a foil composed of three unique items from three different triplets. Target triplets and foils were presented equally often during test. The proportion of correct trials was 67,8% for the syllable condition and between 66,5% and 71,4% for the other conditions, statistically above chance. This study confirms that SL is still under use in late adulthood for several modalities. The following logical step is to investigate how speech SL develops across the lifespan by comparing the performance of different age groups in a same task.

Even if comparative studies between younger and older adults are lacking in speech SL, these are plentiful in implicit learning, a paradigm quite similar to SL. Reviewing the results of these studies of implicit learning may give us an idea of what to expect in a speech statistical learning comparative study of this sort.

1.5 Extrapolating from the results of implicit learning

Implicit learning (IL) research as emerged as a cognitive scientific field almost half a century ago (Reber 1967) with the aim of understanding how information is incidentally (i.e. without awareness) acquired from the environment, as opposed to explicit learning, that requires awareness and the acquired information can be expressed verbally. Several studies on implicit learning have adopted the artificial grammar learning (AGL) paradigm developed by Reber (1967): after exposition to several strings of letters generated by a Markov-chain finite-state automata, participants are then informed that those strings followed a rule. They are then shown, in each trial, two strings, one that obeys to the rules of the grammar and another that does not, and asked to choose the one that does, effectively performing a two alternative forced-choice task. Even though the participants are unaware of these rules during the familiarization phase, the literature show that they consistently perform above chance thus suggesting that they were able to extract the rules (Reber 1967; Perruchet 2008). It has been argued that statistical learning and implicit learning have the same underlying mechanisms of learning and pursue similar goals (Perruchet & Pacton 2006), so much so that some authors fuse both paradigms under a common terminology: implicit statistical learning (Conway & Christiansen 2005). While ALL and AGL might indeed be very similar, other tasks used in IL, such as Serial Reaction Time (SRT) task differ from speech SL in regards to the evaluation of participants' performance. In each trial of SRT (usually) a visual target appears in one of several possible locations in a screen and the participants must press the button that corresponds to that location. Unbeknownst to participants, target location follows a statistical pattern that is repeated through several blocks and the latencies of the motor responses are assessed. While AGL (and ALL) performance is evaluated after practice (offline period) with a 2AFC task, STR performance is directly assessed through the reaction times (RT) during the familiarization phase (online period).

IL research on older adults started in the early 1990s and continued to this day, spanning different tasks, stimuli, and modalities (for a review, see Rieckmann & Backman 2009). Howard and Howard (1989, 1992) were the first to compare young and older in a SRT task, demonstrating an increased typing speed for repeating sequences compared to random sequences for both groups, although overall RT was longer in the older group. Negash et al. (2003) used a paradigm similar to SRT to examine implicit learning: they presented higher-order probabilistic sequences (i.e. the statistical relationship between stimuli occurs among nonadjacent items) consisting of letters, in the center of the screen, each matched with a button, to young and old participants. Both groups showed learning although learning onset occurred later in the older group. In AGL, younger and older adults are equally successful at classifying letter strings as familiar based on an invariant feature (Howard et al. 2008), and at classifying letter strings based on a grammatical rule under implicit conditions (Davis et al. 1998; Meulemans & van der Linden 1997). These results suggest that age deficits appear mainly with higher-order sequences and that older adults show a general deficit in IL rather than a domain-specific deficit. They are somewhat surprising since cognitive aging leads to a decline of several cognitive abilities (Lezak et al., 2012).

1.6 Cognitive Aging

Cognitive abilities change from young to late adulthood (Salthouse, 2010). Aging brings a decline in structure and function of several brain regions (for a review see Dennis and Cabeza, 2008; Raz, 2005), affecting their relative contribution to task performance (for a review see Park and Reuter-Lorenz, 2009). Li & Lindenberger (1999) have proposed that structural changes increase neural noise[2]. Noise is responsible for variability in human performance (Slifkin & Newell 1998; Szalma & Hancock 2011) and older adults present more interindividual variability than young people for several tasks (Lindenberger & Oertzen 2006). Increase of noise might also be responsible for the slowness of processing speed characteristic

[2] The term “noise” is used in information theory for anything that interferes with the transfer of information.

of cognitive aging, spanning most tasks (Salthouse 1996). A particularly big decline can be observed in the frontal lobes (Raz et al. 2005), at least partially responsible for working memory (Braver and West, 2008; Borella et al., 2008) and attention (Zanto & Gazzaley, 2014), two cognitive systems have been associated with speech SL (Toro et al., 2005; Kersten & Earles, 2001).

1.7 Cognitive systems involved in speech statistical learning

Attention and working memory are important for selecting and maintaining in memory relevant stimuli, diminishing interference. Here we explore in more detail the importance of these cognitive systems for SL.

1.7.1 Attention

Attention is a cognitive system that enables the selection of, and maintenance of sustained focus on, relevant information for behavior (Pashler, 1998). Due to the incidental nature of SL, attention was thought not to be of great importance but recent studies employing divided-attention paradigms have supported the conclusion that there is an implicit attentional bias for regularities during SL (Zhao et al., 2013) and that when attention is compromised SL is degraded (e.g., Jiang & Chun, 2001; Baker et al., 2004; Turke-Browne et al, 2005). Perhaps the most indicative study is that of Toro et al. (2005) who evaluated young adults capacity for speech segmentation in an artificial language learning paradigm under two conditions: low-attentional load (i.e. passive listening: their task was to simply listen to the AL) and high-attention load (i.e. their task was to press a button whenever they would hear a change in pitch). While the passive listening group performed significantly above chance in the two alternative forced-choice test, the high-attention load group did not. These experiments show that performance under a dual task condition is degraded compared to a single task, thus indicating that attention is involved in SL. Fernandes (2007, Experiment 5) further accessed the impact of familiarization time (i.e., between-participants familiarization phases of 7min, 14min, 21min) and attention load (i.e., weak and strong cognitive noise) in young adults capacity for statistical speech segmentation of words with different salience (low- and high-TP-words). Interestingly, her results show that in the high-attention load condition participants succeed to extract high-TP-words independently of the familiarization time, but only succeeded to extract low-TP-words in the 7min familiarization. In the low-attentional load condition participants extracted both words above chance, but extraction performance of high-TP-words was only higher than low-TP-words in the shortest familiarization time (Fernandes, 2007, Experiment 5). Some researchers proposed that the attention load could disrupt other mechanisms necessary for learning, such as working memory (Frensch & Miner, 1994).

1.7.2 Working Memory

Working memory (WM) is the capacity to maintain and manipulate information (Baddeley, 1986). WM appears to be modality specific, so we will only mention studies of verbal WM. The performance of young and older participants was correlated with verbal WM in a SRT, corroborating its importance for implicit learning (Bo et al., 2012). In speech SL it also holds considerable importance according to Newport's (1990) *Less is More hypothesis,* that maintains that a restricted working memory is an advantage to language acquisition because it leads to information loss which can be beneficial in highlighting simple contingencies in the environment. The hypothesis is mainly based in the assumption that it is the children's constrained WM, relative to that of young adults (Gathercole, Willis, Baddeley & Emslie, 1994), that enables an easier language acquisition. Evidence for this hypothesis comes from computational simulations (e.g. Elman 1993) and empirical findings (e.g. Kareev et al, 1997; Kersten & Earles, 2001). However, the results of these studies have been disputed (for a review see Rohde & Plaut 2002) and other studies could not hold the hypothesis (Ludden & Gupta 2000).

1.7.3 The impact of aging on both systems

Multiple systems are impaired by cognitive aging, including attention and WM (Craik & Salthouse, 2000). There is a significant correlation in the degree of early suppression and subsequent WM performance in young and older adults (Clapp & Gazzaley, 2010). Older adults fail to ignore distracting information (Gazzaley et al., 2005), which may lead to WM deficits (Gazzaley & D'Esposito, 2007). This inhibitory deficit has been suggested as a source for the broad spectrum of cognitive deficits that occur in older adults (Hasher and Zacks, 1988). According to Campbell et al (2010), this deficit can lead not only to an *hyper-encoding*, in which irrelevant information is encoded along with relevant information, but also to an *hyper-binding,* in which irrelevant information is associated to relevant information. To describe in further detail the impact of aging on these systems and in SL we must look into to the brain at the cellular level.

1.8 Neural Correlates of speech statistical learning

To understand the neural correlates of SL we will briefly introduce the neuron, the elementary computational unit of the brain, and the structures of the brain known to be related to associations between events. We then proceed to review the studies of functional Magnetic Resonance Imaging (fMRI) that have been localizing the brain regions responsible for SL.

1.8.1 The plasticity of neurons

All brain regions are made up of two types of cells: neurons, electrically excitable cells, and glial cells that have the role of supporting the neurons. A typical neuron has a cell body (soma), dendrites and an axon that connects the cell body to the dendrites of the next neuron. The axons of some neurons have a fatty sheath, myelin, that insulates them and increase the speed with which the signals can be transmitted. As a simple terminology, it is usual to call *white matter* to areas of the brain populated mainly by the axons of neurons, due to the white myelin, and *grey matter* to areas of brain populated mainly by cell bodies, due to the darker coloration of the soma. The point of contact between an axon terminal part of a neuron and a dendrite of another neuron is a synapse. Synapses are extremely dynamic: their strength (or weight) can be increased or decreased, depending on the stimulation they get. This capacity, termed synaptic plasticity, is thought to be important to learning and memory (Hebb, 1949; Gruart et al., 2006), correlating with the degree of perceptual learning in auditory tasks in rats (Polley et al., 2006). Synaptic plasticity is divided in short-term and long-term depending on the duration of its effects. Short-term plasticity might be in the base of selective attention and working memory (Jääskeläinen et al 2011) and long-term plasticity in the base of learning and long-term memory. Synaptic plasticity is regulated by signaling pathways, dependent on gene expression and protein synthesis. In general, genes involved in the regulation of synaptic plasticity (for a review see Bishop et al., 2010) are down-regulated (i.e. have a lower expression) during aging.

1.8.2 Associative structures of the brain

Sensory processing is accomplished in the cortex, the convoluted sheet of layered neurons that appears on surface views of the brain. Different stimuli modalities are processed in different locals of the cortex; the auditory cortex is located in the temporal lobe. The medial temporal lobe (MTL), which consists of the hippocampal region and the adjacent perirhinal, entorhinal, and parahippocampal cortices, is involved in forming rapid associations between events unrelated temporally or spatially (Eichenbaum, 2000). Another area that specialized in the stimulus-response binding, based on statistical likelihoods of stimulus occurrence over time, is the striatum, part of the basal ganglia. There seems to be a “competition” between these two systems: recruitment of the striatal system is accompanied by a disengagement of the MTL (Poldrack and Packard, 2003; Dennis & Cabeza 2010). The dynamics of this interaction change as learning progresses: Durrant et al. (2012) exposed participants to a statistically structures sequence of auditory tones and discovered that there is a greater connectivity between striatum and MTL and ventromedial prefrontal cortex after 30 minutes of exposure and a greater connectivity between striatum and planum temporale, an area of the MTL, after 24 hours. Thus, MTL regions predominate in early training while striatal regions would become increasingly important as training progresses.

1.8.3 Studies of functional magnetic resonance imaging

Cerebral blood flow is correlated to neural activity. Through the measurement of changes in flow, fMRI enables the detection of brain areas responsible for different stimuli.

McNealy et al. (2006) were the first to study speech segmentation using fMRI. In the familiarization phase of this study 27 young participants listened to three (approximately 2,5 minutes each) counterbalanced streams of nonsense speech: unstressed and stressed (initial syllable) artificial languages with 12 trissilabic words and a stream composed of random syllables. ALs syllable TPs were set to 1 for syllables within a word and 0.33 for syllables across boundaries. When AL conditions were contrasted with the random syllables condition, there was an increase of activity over time in the superior temporal gyrus and transverse temporal gyrus, extending into the supramarginal gyrus in the left hemisphere and a decrease of activity in the ventromedial prefrontal cortex. When more liberal thresholds were applied, these increases extended to the left superior temporal gyrus and bilateral inferior parietal lobule and the decreases to the left medial temporal girus, inferior parietal lobule, right, middle and medial frontal gyri, anterior cingulate, and left putamen. Fifteen of the 27 participants completed a second fMRI task where they listened to blocks of words and part-words (in the unstressed version) from the unstressed and stressed AL and non-words from the other stream. A statistical contrast comparing BOLD signal associated with listening to words versus nonwords (W - NW) revealed significantly greater activity for the words in left inferior frontal gyrus and middle frontal gyrus. The same pattern was observed when comparing words versus part-words (W - PW) and part-words versus non-words (PW - NW). PW - NW also revealed greater activity in superior temporal lobe. Further fMRI studies had similar outcomes (Cunnilera et al., 2009; Karuza et al., 2013).

Schapiro et al (2013) report a case study of a patient with complete bilateral hippocampal loss and some additional damage to surrounding medial temporal lobe cortex and left anterior temporal lobe. Compared with healthy controls[3], with a performance statistically above chance, her performance remained equal or lower than chance, reinforcing the importance of the medial-temporal lobe for SL.

Aging is associated with brain size and weight decrease from early to late adulthood (e.g., Kemper 1994), although not uniform over its several areas (Raz et al., 2005). Gray matter declines in a linear fashion from the 20s to the 50s with a later plateau of the age trend, whereas the volume of the white matter increases from young adulthood to middle age, and then declines with aging (Sullivan et al., 2004). Raz et al. (2005) performed a longitudinal study of five years with participants ranging from 20 to 77 years and found that prefrontal cortices shrinkage is of 0.91% per annum, temporal cortices shrinkage is of 0.59% per annum and striatum shrinkage is of 0,75% per annum. This volume loss might be inflicted by shrinkage or loss of neurons (Kril et al., 2004), loss of dendritic arborization (Jacobs et al., 1997), loss of intralaminar myelin (Courchesne et al., 2000), among other factors.

The decline in the structural integrity of white matter tracts (i.e. demyelination; Rabbitt et al. 2007a ; Turken et al. 2008) or volume loss (Rabbitt et al. 2007b) could result in slowing of information processing (Kemper, 1994; O'Sullivan et al., 2001), associated with old age.

1.9 Other theories as the solution to the segmentation problem

SL seems to be subject to a common tendency toward "reduction of uncertainty" (Gibson, 1991), which supports the idea of the mind essentially as an associative engine where cognitive processing is achieved through statistical computations similar to those implemented in connectionist neural networks (e.g., Elman, 1990; Bates & Elman, 1996). Other authors

[3] Please see the section "Speech statistical learning across lifespan" for more details about the study and controls.

pursue nativist theories that postulate that cognitive processing is achieved through formal symbolic operations (e.g., Chomsky, 1980; Fodor, 1975). Yang (2004) proposed the *Unique Stress Constraint* as the innate rule to word segmentation: a word has only one primary stress and if two stressed syllables are adjacent a word boundary must be between them. However, many languages, such as Japanese, do not have word-level stress. Attempts to identify an innate rule to speech segmentation will most certainly fail due the enormous diversity of phonological features in natural languages (see Mielke 2008 for a survey of meaning-bearing sounds on 628 language varieties). Endress & Bonatti (2007) try to reconcile the connectionist and the inatist views with the *More than One Mechanism Hypothesis*: rule-based learning would assist in the extraction of words that follow these rules and statistical learning would then assist in the extraction of the "deviant" words.

Alternatively, some authors (e.g. Perruchet and Vinter, 1998) believe that speech segmentation is achieved through the concatenation of perceptually relevant initial stimuli primitives into bigger chunks that are stored in memory and then compete with each other by attentional resources. This approach is known as *chunk-based approach* or *clustering approach* (Goodsitt et al., 1993) since it is the clustering of initial primitives with strong associations between them that enable the discovery of the words in the speech stream. Even though learning effects based on nonadjacent TPs (e.g., Onnis et al., 2005) seem to challenge the clustering approach, the computational model PARSER by Perruchet and Vinter (1998) has had some success explaining behavioral results (Perruchet & Vincent 2014). When the performances of PARSER and SRN are compared to that of human learners, PARSER seems to have a better fit (Giroux & Rey, 2009; Frank et al., 2010). Because there has never been a comparison of older adults' data between models, it would be of interest to do these same comparisons in the present work.

1.10 Hypotheses and Objectives of this experiment

As described in the literature revision, aging lead to the impairment of several cognitive systems important for the normal functioning of speech SL. Being so, if compared to younger adults for a same speech SL task, older adults should perform worse. Implicit learning studies on artificial grammar learning show indeed this decline in performance in the older group when compared to the young one. Since both paradigms share many similarities, we hypothesized that a similar result is to be expected in an artificial language learning paradigm.

The objective of this study was to access the performance of older adults (average age 62 yo) in comparison with young adults (average age 20 yo) in an artificial language learning paradigm. We used the artificial language of Fernandes et al (2007) that, by comprising different values of transitional probabilities between its syllables, enables a fine-grained analysis of the importance of these statistics to the segmentation of speech.

2. Method

2.1 Participants

Sixty-eight monolingual European-Portuguese speakers participated voluntarily in this study, which was approved by the ethical committee of Faculdade de Psicologia of Universidade de Lisboa. Participants from the young group (age range: 18-25 years old) were undergraduate students of Psychology from the Universidade do Porto and Universidade de Lisboa, Portugal, and received course credits for their participation. The older group (age range: 54-70) were students from senior universities of Lisbon and Porto and received qualitative information on memory assessment as an incentive for participation. Fourteen participants (6 young and 8 older adults) were excluded based on six criteria: (i) performance below the cutoff score in the Portuguese version of the *Montreal Cognitive Assessment, MOCA* (i.e., below 26; Freitas et al., 2013) (1 young and 3 older adults); (ii) hearing impairment (2 older adults); (iii) taking medication for psychiatric disorders (2 young adults); (iv) having a neurological disorder (1 older adult), (v) failing in understanding the instructions of the procedure (1 young adult); (vi) unsuitable age (1 older adult).

The final sample comprised 27 participants (18-23 years, *M* age = 19.63, SD = 1.47; 24 female) of the young group and 27 participants (54-68 years, *M* age = 60.67, SD = 3.92; 22 female) of the older group.

To gather information on cognitive functioning of the participants and to ensure that none of the older adults had any cognitive impairment, participants performed five ancillary tests. Table 1 presents the average performance of young and older adults in the ancillary measures and the comparison between groups.

Ancillary measure (Maximum value)	**Young** ***M (SD)***	**Older** ***M (SD)***	**Comparison (t-test)**
Vocabulary Subtest** (60 points)	**41,48 (8,67)**	**49,70 (6,94)**	t(52) = -3,85, p = 0,0003
MoCA* (30 points)	28,37 (1,33)	27,30 (1,64)	t(52) = 2,64 p = 0,011
Digit Span Subtest Direct (18 points)	7,04 (2,12)	7,44 (2,12)	t(52) = -0,71 p = 0,483
Digit Span Subtest Inverse (18 points)	5,33 (1,94)	5,59 (2,26)	t(52) = -0,45 p = 0,653
Digit Span Subtest Sum (36 points)	12,37 (3,66)	13,04 (3,99)	t(52) = -0,64 p = 0,525
Stroop Word Card Task** (item/s)	**2,14 (0,29)**	**1,89 (0,37)**	t(50) = 2,71 p = 0,009
Stroop Color Card Task (item/s)	1,48 (0,38)	1,38 (0,22)	t(50) = 1,10 p = 0,276
Stroop Color-Word Card Task** (item/s)	**1,06 (0,16)**	**0,91 (0,16)**	t(50) = 3,37 p = 0,001
AL Syllable Perception (100%)	0,94 (0,04)	0,93 (0,06)	t(52) = 0,53 p = 0,600
AL Syllable Production (100%)	0,88 (0,10)	0,88 (0,14)	t(52) = 0,04 p = 0,972

Table 1 - Performance and comparison of both age groups in the five ancillary tests. N = 27 for each age group in every test except for the three Stroop Card Tasks: young group N = 27, older group N = 25. * Comparison is significant at the 0.05 level (2-tailed)

The vocabulary subtest of WAIS-III (Wechsler, 1997) is a test of accumulated verbal knowledge, involving long-term memory, concept formation and language development. It is highly resistant to neurological deficits and psychological disturbance, being well maintained during aging (Lezak, 2012) and even showing moderate increases until age 60 (Wechler, 1997). Participants were asked the definition or a synonym of each word presented (maximum of 30 words). Each answer is rated as 0, 1 or 2 points with maximum score of 60 points. After three consecutive errors the test ends. The raw score was converted to a scaled score according to participants' age group. Older participants performed significantly better than young participants (Table 1). This was expected since verbal accumulated knowledge increases during lifetime (Wechler, 1997).

MOCA (REF; Portuguese version: Freitas et al., 2013) is a cognitive screening instrument for detection of mild cognitive impairment and dementia in older adults. . It comprises seven subtests: visuospatial/executive (5 points), naming (3 points), memory (5 points for delayed recall), attention (6 points), language (3 points), abstraction (2 points), and orientation (6 points), giving a total of 30 points. A global score below the cut-off score (of 26) is suggestive of cognitive impairment In the present study, as shown in Tabl1, older participants performed significantly worse than young participants, even though all had a performance above the cut-off score (Table 1).

The digit span subtest of WAIS-III (Wechsler, 1997) is a test of short-term memory and attention. Participants are asked to repeat the sequence of digits (of a minimum 3 digits, with increasing length) presented orally by the experiment in forward (in the first list) and in or backwards (i.e., in reverse order, in the second list). For each sequence correctly produced one point is given (maximum: 18 points). The list is terminated if the participant present two incorrect responses in a row. Participants who are easily distractible could have problems performing Digits Forward, while participants might perform badly in Digits Backward if they have short-term memory problems. While Digits Forward is resistant to cognitive deterioration, Digits Backward is quite sensitive to it (Lezak, 2012). High global scores suggest a good auditory short-term memory and good attention and low scores suggest difficulty in concentration or poor short-term memory. There was no significant difference between young and older participants (Table 1).

The Stroop test (Stroop, 1935) is an executive test, of selective attention and inhibition control. Participants performed the three tasks of the Portuguese Stroop test (Castro, Cunha & Martins, 2000): the word card task, the color card task and the color-word card task, each comprised 112 items displayed in a 4 x 28 matrix in a A4 sheet. Each task was limited to 120 s.

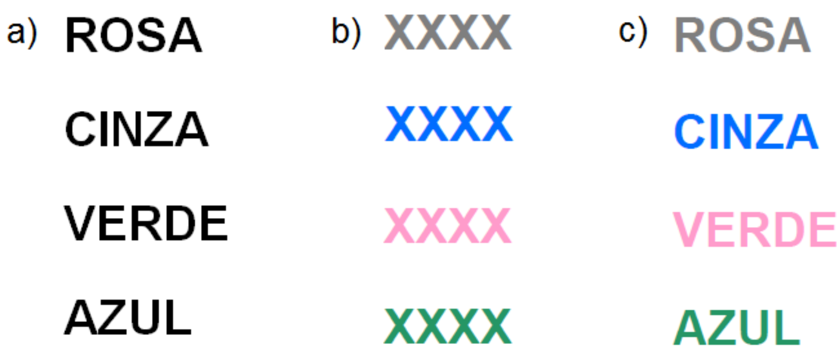


Figure 2 - Examples of the three tasks of the Stroop test: word card task (a), color card task (b) and color-word card task (c).

In the word card task the sheet comprises 112 Portuguese words (ROSA, CINZA, VERDE, AZUL; in English PINK, GREY, GREEN and BLUE) printed in black (Figure 2A). Participants were asked to read the stimuli from the top-left in column until the bottom-right and time was registered. If the 120 s elapsed, the number of items performed correctly was considered. In the color card task, the sheet comprised XXXX printed in one of four colors (pink, grey, green and blue; Figure 2B), and participants were asked to name the stimuli from the top-left in column until the bottom-right. Finally, in the color-word card task, the sheet comprised the words used in the word card task, but which are presented with an incongruent ink (Figure 2C); participants were asked to name the color of the ink (and not to read the words) with the same procedure as in the previous tasks. Performance in each task was calculated by dividing the number of items by the time spent in the task (maximum of 120 s). Participants usually take longer to complete the color-word card task than to complete the color card task. This discrepancy was termed *Stroop interference* and arises because in the latter task the automation of reading is faster than the identification of the color, which means participants need to inhibit the urge to read the word and must instead say the color. The Stroop interference was computed as the difference in performance between the color naming task and the color-word task. Data of two older participants was missing: one because he was daltonic, the other due to technical problems. There was a significant difference between young and older participants in the word card test and the color-word card test, but not in the color card test (Table 1). This difference has been noted in the literature (e.g. West & Alain., 2000) and is attributed to cognitive aging of the older group, either due to general slowing effects (Verhaeghen & De Meersman, 1998), inhibition deficits (Hasher & Zacks, 1988) or dysfunction of frontal lobes (Stuss et al., 1994; West, 1996).

Immediately before the experimental task participants performed an AL syllable perception and production tests in order to assess participants' acuity and to ensure that any difference between the two groups in ALL would not be due to low-level, perceptual problems of the older group, such as hearing loss. Stimuli presentation and data collection were controlled by E-prime 1.1 (Schneider et al., 2002; Psychology Software Tools, Pittsburgh, PA). The syllables used in these two tests (/bɐ/, /dɐ/, /tɐ/, /zɐ/) are minimal pairs (i.e. differ in only one phonological element), and were synthesized using the text-to-speech MBROLA software (Dutoit et al., 1996), with a European-Portuguese female diphone database (available at www.tcts.fpms.ac.be/synthesis/mbrola.html) at 22.05 kHz. In the perception test, participants performed a same-different comparison task for 48 trials. In each trial, two sequential stimuli, each chosen from the pool of four, were presented with a 500 ms silence between them. In the production test, in each trial participants were presented with one of these four stimuli and were asked to shadow it, for a total of 32 trials. Order of trials was randomized for each participant in both tests. Participants' responses were registered by the experimenter on the computer's keyboard. Responses did not differ significantly between age groups in neither perception nor production (Figure 2), ensuring therefore that the older participants had the same level of perceptual acuity than younger adults.

2.2 Material and Procedure

The AL adopted here was that of Fernandes et al. (2007, 2010). The phonological repertoire of the AL comprised five consonant phonemes (/b, /p/, /k/, /l/, /f/) and three vowel phonemes (/ɐ/, /i/, /u/) arranged in 12 syllables in order to create six TP-words. All TP-words occurred with the same absolute frequency in the AL stream during the familiarization phase, but they varied on average TP between syllables. Half of the TP-words had an average TP of .52 and were called *low-TP-words*; the others had an average TP of .83 and were called *high-TP-words* (see Table 2).

Word	TP	TP group
/bukɐlɐ/	.49	
/kɐfubi/	.50	low
/bɐbuku/	.58	
/fufibu/	.75	
/kilɐbu/	.75	high
/lufɐbɐ/	1	

Table 2 - Words of Fernandes et al. (2007) language, with respective TPs percentage and the TP group they belong to.

For the familiarization phase, three blocks of 7 min each were synthesized using the text-to-speech MBROLA software (Dutoit et al., 1996), , with a speech rate of ~270 syllables per min and following the same procedure used in the AL syllable ancillary tests. Each block was created by concatenating 105 tokens of each TP-word type (total of 630 tokens, 1890 syllables) with the only criterion that no TP-word occurred twice in a row. The AL blocks were presented with Windows Media Player, at a comfortable level through *Sennheiser HD 280* headphones. Participants were instructed to listen to a new language, without grammar, made of meaningless words. Their task was to find out those words. No information about the structure, phonology, length, or number of the words was given. Participants were further informed that they would hear the AL in three blocks, separated by a short resting period, and would be tested afterwards on their knowledge of these words.

The familiarization phase was immediately followed by a two-alternative forced choice test, with stimuli presentation and data collection controlled by E-Prime. Six *part-words* were selected for the two-alternative forced-choice test. They comprised one syllable of one TP-word and two syllables of another TP-word that had occurred adjacently in the AL stream during the familiarization phase, either the two last syllables of a TP-words and the first syllable of another one or the last syllable of a TP-word and the first two syllables of another one (see Table 3).

Part-Word	TP
/lɐbu#fu/	.31
/ku#bukɐ/	.34
/kɐlɐ#fu/	.36
/fibu#lu/	.50
/bɐ#kilɐ/	.52
/bi#lufɐ/	.55

Table 3 - Part-words of the AL used in the two-alternative forced choice test, with respective TPs values (these are not further analyzed and are displayed just for reference). # corresponds to a boundary between TP-words.

In each trial, participants were presented with two trisyllabic stimuli, a TP-word and a part-word, with a counterbalanced order. Each TP-word was paired exhaustively with each part–word, in a total of 36 trials. Participants were asked to decide which one of the two stimuli was a word of the language heard in the familiarization phase. Response was given through key pressing: key “1”, right-index finger, or key “2”, left-index finger, if they decided that the word was the first or the second stimuli, respectively. After their response, or if no response was given after 10 s, a new trial began. Participants were asked to provide an answer even if they were not sure about it, but accuracy was emphasized. The test began with four practice trials to enable practice with key pressing and to ensure that participants understood the task. In the practice trials, animal and environment sounds were used as stimuli and participants were asked to decide which corresponded to an animal sound. Feedback was provided only for these trials. When participants’ performance was at chance in these practice trials (i.e., less or equal to 4 correct responses), these trials were repeated after ensuring that participants understood the task. Next, they performed the experimental trials (on the AL) without any feedback. Order of presentation of the test trials was randomized for each participant and order of presentation of the stimuli within-trial was counterbalanced within each group.

3. Results and Discussion

Response accuracy in the two-alternative forced choice test was computed for each participant by type of TP-word (low TP vs. high TP), and analyzed with a mixed 2 x 2 ANOVA with age group (young vs. older) as a between-subjects factor and TP-word type (low-TP-word vs. high-TP-word) as within-subject factors.

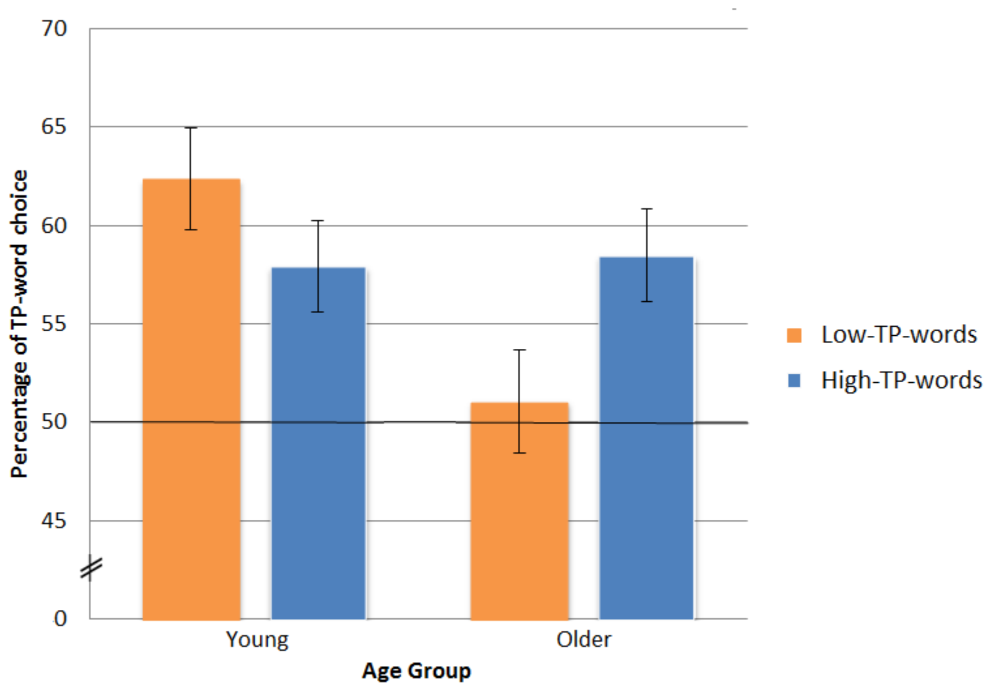


Figure 3 - Percentage of each TP-word group chosen by young and older participants in the two-alternative forced choice test. Error bars represent standard error of the mean.

The main effect of age group was significant, $F(1, 53) = 5.83$, $p = .019$, suggesting that overall older adults performed worse than young adults in the ALL (Figure 3). More importantly, the effect of age was modulated by TP-word type, $F(1, 53) = 4.67$, $p = .035$ (main effect of TP-word type: $F(1, 53) = 0.36$, p = .554). Indeed for high-TP-words the older group presented the same level of learning as the young group, t(52) = 0.17, $p = .867$, and both were able to select the high-TP words significantly above the chance level: the young group, $t(26) = 4.82$, $p < .001$, the older group, $t(26) = 3.34$, $p = .003$. In contrast, for low-TP-words, the young group outperformed the older group, $t(52) = 3.08$, $p = .003$. Whereas the young group was able to correctly select the low-TP-words significantly above chance, $t(26) = 4.82$, $p < .001$, with a performance similar to that for high-TP-words, t(52) = 1.13, $p = .265$, the older group was unable to correctly select the low-TP-words, t(52) = 0.41, $p = .685$, with a performance almost statistical significantly different from that of high-TP-words, t(52) = -1.94, $p = .058$.

Since there was a significant difference in MoCA performance between age groups (Table 1), we re-examined group differences in the two-alternative forced-choice test performance by doing an analysis of covariance (ANCOVA) with MoCA results as covariate. The ANCOVA revealed that the interaction between TP-word type and age group was still significant, $F(1, 52) = 3.95$, $p = .052$. This suggests that the difference in performance between age groups in the MoCA test is not the cause of the difference in performance between age groups in the two-alternative forced-choice test.

Low-TP-words have lesser salient TP values than high-TP-words which make them more attentional demanding to find in the speech stream. Plus, their TP values are similar to those of some part-words, further difficulting their extraction in comparison with high-TP-words. Older adults' attentional deficits compromise the effectiveness of working memory to maintain in memory relevant stimuli and associations, leading to an increase of interference.

Aging also increases cognitive noise, due to declines in white (Turken et al. 2008) and grey matter (Raz et al, 2005), as well as a down-regulation of genes involved in the regulation of synaptic plasticity (Bishop et al., 2010). Fernandes et al., (2010) examined the performance of young adults under different attention load conditions (i.e. low and high), showing that while under weak cognitive noise both low- and high-TP-words are successfully chosen above chance in the 2AFC, under high cognitive noise only high-TP-words are successfully extracted from the AL. Natural cognitive noise of older adults might have the same effect as induces cognitive noice in young adults. In a previous study Fernandes (2007, Experiment 5) examined the performance of young adults under not only different attention load conditions (i.e. low and high) but also different AL exposition times (i.e. 7, 14 and 21 min). The performance of younger adults' performance at 14 min in the high load condition of their study seems strikingly similar to the older adults' performance at 21 min in this study. This raises the question if older adults could reach the performance of younger adults at 21 min. given enough time.

Correlational analyses were used to examine the relationship between participant's cognitive abilities and performance in the ALL. No statistically significant correlations were found for the young group (Table 4). In the old group, the high-TP-word score and average score between low- and high-TP-words were inversely related to WAIS-III Digit Span Inverse score (Table 5). The low-TP-word score and the average were inversely related to WAIS-III Digit Span Sum score and Stroop Color-Word Card Task (Table 5). The Digit Span Sum score accesses WM, which is thought to be important for SL, so it is intriguing that a lower WM would increase the performance in the task. An explanation might lie in the *Less is More hypothesis* of Newport (1990), that maintains that a more limited working memory span may provide an advantage for statistical SL by highlighting simple contingencies in the environment.

		Vocabulary Subtest	Moca	Digit Span Subtest Direct	Digit Span Subtest Inverse	Digit Span Subtest Sum	AL Syllable Perception	AL Syllable Production	Stroop Word Card Task	Stroop Color Card Task	Stroop Color-Word Card Task	Stroop Interfer.
Low-TP-Words	Pearson Correlation	,143	-,006	-,187	-,258	-,245	-,179	-,018	-,132	,050	-,312	,055
	Sig. (2-tailed)	,476	,977	,351	,195	,218	,372	,929	,513	,803	,113	,784
High-TP-Words	Pearson Correlation	,126	,303	-,096	,178	,039	-,027	,207	-,184	-,002	,065	,130
	Sig. (2-tailed)	,531	,124	,634	,375	,848	,893	,300	,357	,992	,748	,519
Mean	Pearson Correlation	,192	,203	-,211	-,080	-,164	-,159	,137	-,235	,037	-,200	,139
	Sig. (2-tailed)	,338	,311	,291	,693	,413	,428	,497	,238	,855	,318	,489

Table 4 - Correlations between ancillary tests performance and TP-word groups performance of the young group. * Correlation is significant at the 0.1 level (1-tailed); ** Correlation is significant at the 0.05 level (2-tailed).

		Vocabulary Subtest	Moca	Digit Span Subtest Direct	Digit Span Subtest Inverse	Digit Span Subtest Sum	AL Syllable Perception	AL Syllable Production	Stroop Word Card Task[1]	Stroop Color Card Task[1]	Stroop Color-Word Card Task[1]	Stroop Interfer.[1]
Low-TP-Words	Pearson Correlation	,227	,092	-,291	-,301	**-,325***	-,115	,065	,057	-,066	**-,363***	,283
	Sig. (2-tailed)	,255	,647	,141	,127	,098	,568	,748	,788	,753	,075	,170
High-TP-Words	Pearson Correlation	-,229	-,166	,019	**-,379****	-,205	-,169	-,033	,209	-,062	-,142	-,071
	Sig. (2-tailed)	,251	,407	,925	,051	,306	,400	,869	,316	,769	,499	,736
Mean	Pearson Correlation	,013	-,036	-,186	**-,451****	**-,354***	-,191	,031	,163	-,092	**-,353***	-,245
	Sig. (2-tailed)	,948	,860	,352	,018	,070	,340	,878	,435	,663	,083	,238

Table 5 - Correlations between ancillary tests performance and TP-word groups performance of the older group. 1 The data of two participants was unavailable and therefore not taken into account for the calculations of the correlations between the three Stroop C ard Tasks and the ALL performance. * Correlation is significant at the 0.1 level (1-tailed); ** Correlation is significant at the 0.5 level (2-tailed)

Computational Experiment

4. Introduction

4.1 Computational Modeling

Computational modeling helps assessing the plausibility of an explanation (e.g., theory) by instantiating it in some quantitative form (e.g., mathematical model). In the cognitive realm, it enables the investigators to test what kind of structural and functional properties of a cognitive process give rise to observed behaviors, promoting further theoretical and practical applications. For successful modeling of cognitive processes a model must satisfy the necessary condition of fitting observed human behavioral data from different tasks evaluating the same process. Goodness of fit has also traditionally been the criteria of selection between competing models in psychology[4]: the model that best fits a particular set of observed data is considered superior. Goodness of fit is dependent upon the number of a model's variables and their range and on the functional form of the model's equation. A model might underfit the data if it has few variables or if they have a small range. On the contrary, if it has too many variables or they have a wide range it might overfit the data. A delicate balance between these parameters is therefore necessary to reach a good fit. Searching for parameter values that produce the best fit provides only post hoc justification for a particular choice of parameters. The ideal situation for a model is to have the same (good) performance on new stimuli while retaining the same parameter values it held in the evaluation of other data. This would provide non post hoc support.

4.2 Model of speech statistical learning

As previously mentioned, there are two different approaches to speech segmentation: the bracketing approach and the clustering approach. The bracketing approach is usually

[4] Even if some authors defend the superiority of other criteria (e.g. Myung, 2000).

implemented by the simple recurrent networks (SRN) model, such as in Elman (1990) or Christiansen, Allen, and Seidenberg (1998), or by a variation of this model[5] (e.g. Frank et al 2010). Models of the clustering approach include the Competitive Chunking (Servan-Schreiber & Anderson, 1990), PARSER (Perruchet & Vintner, 1998), the clustering algorithm of Swingley (2005) and MDLChunker (Robinet et al., 2011). Of these, PARSER has been the most widely used.

4.2.1 Bracketing Approach - Simple Recurrent Networks

The simple recurrent network (SNR), a type of artificial neural network (ANN) proposed by Elman (1990), and also known after his name (i.e. Elman Network), is one of the most used models in the statistical learning literature (Perruchet & Peereman, 2004). ANNs are computational architectures inspired by biological brains (e.g., McClelland et al., 1986) and commonly called connectionist systems.

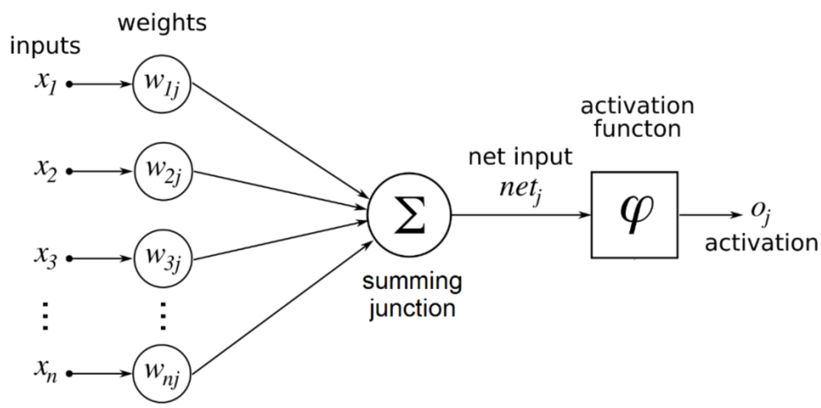


Figure 4 - Elements of an ANN neuron. Adapted from Wikimedia Commons.

The fundamental information-processing unit of ANNs is the node or neuron, which has three basic elements: a range of synapses with associated weights, a summing junction to sum the input signals of the synapses and an activation function for limiting the amplitude of its output (Figure 4).

[5] A recent exception is TRACX, a feedforward network by French et al. (2011)

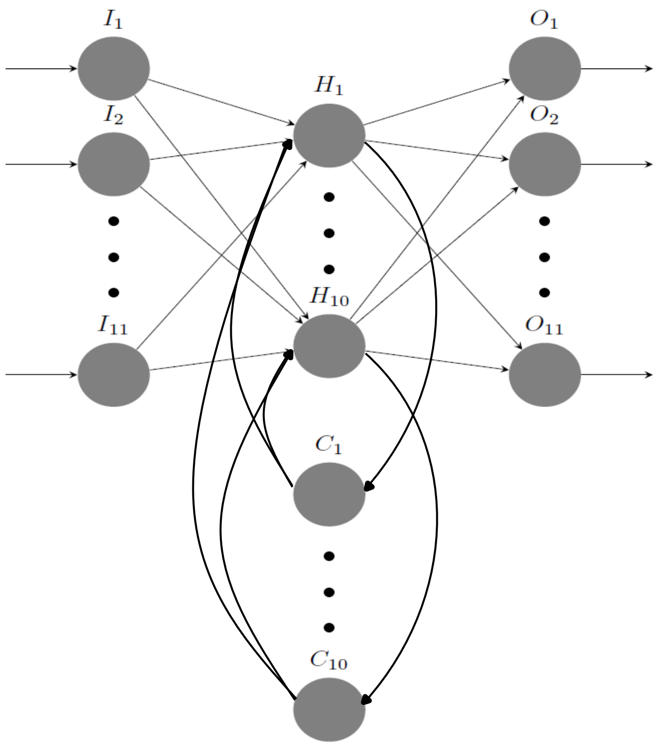


Figure 5 - Example of a SRN with 11 input nodes, 10 hidden nodes, 10 context nodes and 10 output nodes. Note that the connections from the hidden layer to the context layer are one to one while from the connections from the context layer to the hidden layer are one to all.

Neuron units can be divided into three categories, usually grouped into layers: input units that receive data from the outside of the network, hidden units whose input and output remain within the network and output units that send data to the exterior of the neural network (for an example see Figure 5).

ANN can be supervised or unsupervised. In the supervised models there is a training phase where pairs of input-output data are given in succession to the network. In statistical learning modeling the data are usually syllables (as is the case of this study), phonemes or other similar linguistic units. For instance, one of these pairs could be the syllable "cog" as input and the syllable "ni" as output - the first two syllables of the word "cognition". The following pair would be "ni" as input and "tion" as output, and so on. However, since neural networks work with numerical data, a number (usually binary) has to be attributed to each syllable. A very simple AL with only three syllables could be coded as “100”, “010” and “001” for example. Back-propagation is the most used algorithm (Rumelhart, Hinton, & Williams, 1986) designed for supervised feedforward networks. Learning is achieved through two passes: a forward pass, where an input is applied to the input nodes and its signal is propagated by the network layer after layer until an output is generated by the output nodes; and a backward pass where the response of the network is subtracted from the given output in order to produce an error signal which is propagated backwards to adjust the weights of the network. The unsupervised models are fed unlabeled data, and the objective of the network is to find regularities in the data. The incidental nature of SL would, in principle, make it a good candidate to modeling with unsupervised models, but until this date there have been no such attempts.

SRN are an example of a popular approach to temporal pattern recognition: constructing a buffer that holds the n most recent elements of the input sequence, turning a temporal recognition problem into a spatial recognition problem, more appropriate to connectionist models. This buffer is known as context layer and it copies the output of the nodes from the hidden layer (at time t), delivering it again to the hidden layer nodes in the next iteration (time t + 1) of the learning process. This recurrent connection originates feedback, enabling a greater computational ability (Siegelmann, 1999) and mimicking biological networks (Freeman, 1975).

Backpropagation can be used to train recurrent networks given they are “unfolded in time” and thus becoming feedforward. This algorithm is known as back-propagation through time (Rumelhart, Hinton, & Williams, 1986) and is often used in SRN.

After training, the network should have a good representation of the transitional probabilities of the language (Table 6).

	BA	BI	BU	FA	FI	FU	KA	KI	KU	LA	LU
BA	0,09	-	0,63	-	-	0,12	0,14	0,01	-	-	-
BI	0,19	-	0,16	-	-	0,28	-	0,33	-	-	0,04
BU	0,13	-	0,12	-	-	0,04	0,34	0,08	0,25	-	0,05
FA	1,00	-	-	-	-	-	-	-	-	-	-
FI	-	-	1,00	-	-	-	-	-	-	-	-
FU	-	0,50	-	-	0,50	-	-	-	-	-	-
KA	-	-	-	-	-	0,50	-	-	-	0,50	-
KI	-	-	-	-	-	-	-	-	-	1,00	-
KU	-	-	0,08	-	-	0,17	0,18	0,15	-	-	0,42
LA	0,07	-	0,50	-	-	0,08	0,09	0,09	-	-	0,17
LU	-	-	-	1,00	-	-	-	-	-	-	-

Figure 6 - Probabilities of a syllable (y axis) being followed by another (x axis) in the AL. The transitions between syllables of TP-words are colored in green, between syllables of part-words in blue, between syllables of TP-words and TP-part-words in grey and

The 2AFC test is simulated by presenting a word or part-word, one syllable at the time, and accessing the activation of the output unit corresponding to the last syllable/phoneme of that word or part-word. The stimulus with the higher activations is the chosen one.

4.2.1.1 Parameters

There are several parameters that have a strong impact in the network performance. In connectionist models, the ideal value of a certain parameter depends on the problem in question. “Standard values” (i.e. values that have proven to be optimal or near optimal for several problems) can be found from previous literature (Table 1), but every new problem requires a new search of the optimal values from the parameter space. The attribution of the adjectives “high” and “low” to values is relative, and can vary wildly from problem to problem.

Symbol	Parameter	Standard Values
l	*Learning rate*: the rate at which weights change under back propagation algorithm	.1
m	*Momentum:* the degree to which weights continue to change in the same direction	.1
a	*Architecture:* the number of hidden nodes	{10, … 20}
w	*Weights*: the value of the connections between nodes	[-1, 1]
e	*Epoch*: Number of times the network is trained	[1, ∞]

Table 6 - SRN Parameters with a small description and respective standard values for studies of speech segmentation (e.g. Christiansen et al. 1998).

Learning rate

It sets the rate at which weights change in the learning of the training algorithm. A "high" value might reach optimization faster but with a more unstable weight space while a "small" value enables a smother trajectory for the weight space but small rate of learning.

Momentum

It adds a fraction *m* of the previous weight update to the current one, preventing the network from converging to a local minimum. Much like in the learning rate, a "high" value may increase the speed of convergence, as well as the instability of the system, while a low value "slows" the convergence but enables a more stable trajectory.

Number of nodes in the hidden layer

Hidden nodes take the data from the input nodes and transform it. Too many neurons may overspecify the system, thus preventing it from properly generalizing. Conversely, too few neurons may prevent the system from properly fitting the input data. Both situations reduce the robustness of the system and are to be avoided.

Weights

Connections between nodes have weights, normally in the range [-1, 1]. They are usually randomly attributed at the beginning of a simulation, but might be fixed.

Epochs

It is the number of times a network is trained with the same data before progressing to the test phase.

4.2.1.2 Simulation of older participants

The effects of aging have not been simulated in the realm of speech SL, but SRN models have been used previously to simulate brain lesions (Hinton & Shallice, 1991). Hinton & Shallice (1991) introduced three procedures to simulate the effect of lesions in an ANN:

(i) setting some of the connections' weights to 0 after training;

(ii) adding noise, uniformly distributed between two values, to the connections' weights;

(iii) removing some of the neurons from the hidden layer.

These changes could, in principle, be used to simulate aging effects as well, if carefully based in neurological evidence. If connection weights are set to 0 is as if the connection between these two nodes was eliminated. If the performance of the system was optimized with that value and suddenly it is changed after testing, an inevitable increase in the noise of the system will be the result. (i) is a good simulation of white mater decline, observed with aging, because while the nodes remain, the communication between them was compromised. (i) and (iii) will inevitably result in noise, but for higher values of noise (ii) can be considered. SRN models of speech SL have not been used to simulate different brain areas. As previously stated, the neural correlates of SL comprise several brain areas that are affected differently by aging effects (Raz et al., 2005). One way to overcome this problem is to average the shrinkage across the different neural correlates. For instance, if we average shrinkage of prefontral cortices, temporal cortices and the striatum according to Raz et al. (2005) is of 0,75% a year (Table 1).

At this rate from the 20s to 60s there would be a shrinkage of 30%. The ideal way of using (iii) in the simulation of age is perhaps to find the optimum value for nodes in the hidden layer in the simulation of young adults and then remove 30% of that value for the simulation of the older adults.

Slowness of processing speed is a characteristic aspect of cognitive aging, perhaps due to noise. By decreasing the value of learning rate and momentum, important parameters for the backpropagation algorithm, we would increase the time necessary for the model to reach an optimum state for the solution of the problem of speech segmentation.

Finally, attention and working memory are also impaired in old age. One way of simulating this impairment would be by skipping some of the input patterns. For example, for the input patterns "BU|KA|LA", presented in succession, we could omit "KA" so that, for the network, in this case, it "BU" would be followed by "LA".

4.1.2 Clustering Approach - PARSER

PARSER is a well-known model of word segmentation (Krogh, Vlach and Johnson, 2012). It was originally developed to demonstrate that the results from the study of Saffran et al. (1996) could be explained by other processes than TP computations but since then it has been used to fit speech segmentation behavioral data (e.g.: Perruchet & Tillmann, 2010, Frank et all 2010), mainly due to overall good performance (Perruchet & Vincent 2014).

In PARSER word segmentation is grounded on two assumptions: perception shapes internal representations, and internal representations guide perception (Perruchet and Vinter, 1998). In other words, perceptually relevant stimuli start by catching the attention of the learner and then these familiar stimuli serve as a guide to learners' further attentional allocation. A stimulus becomes more relevant by repeatedly appearing in the environment. Stimuli comprising words appear more often together in speech than those comprising between-word, so it follows that speech segmentation may simply rely on mechanisms of memory and associative learning, without the need of computing TPs.

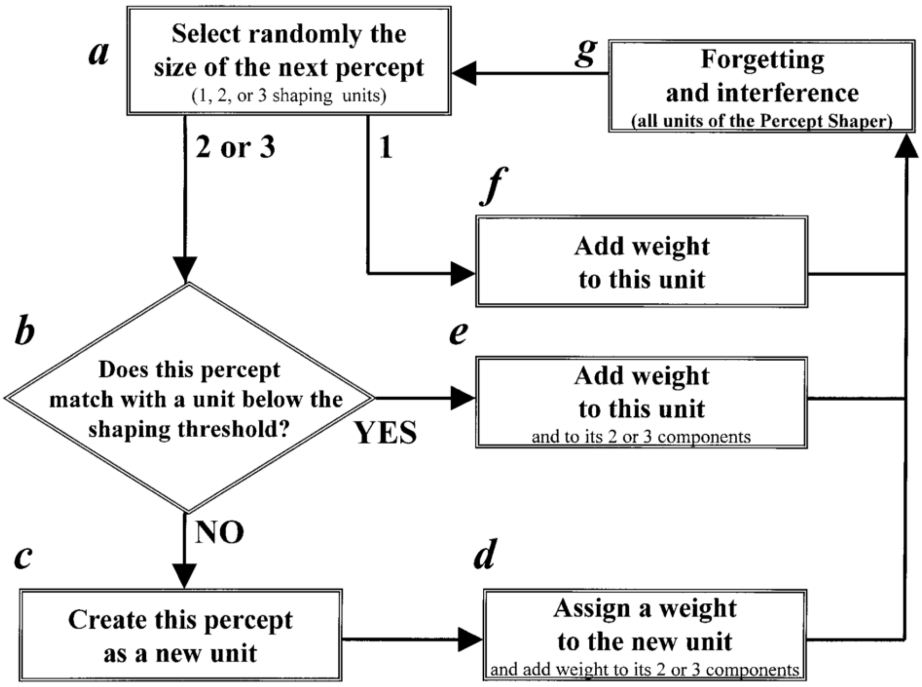


Figure 7: Diagram of PARSER's algorithm for percept (chunk) creation. Original from Perruchet & Vinter (1998).

The model works by randomly picking one percept/chunk with a size from one to three shaping units[6] (Figure 5a; Parameter 1 of Table 2) at each iteration. If the percept is already in the percept shaper (the memory of the model) its weight is increased, otherwise it is added until a total of 20 percepts. Usually in the beginning of the simulation the shaping units are the syllables or phonemes of the AL, but because every percept in the percept shaper becomes a possible shaping unit, after some iterations shaping units can be several syllables/phonemes long.

[6] Miller (1956) proposed that people have a working memory capable of seven chunks. More recent studies suggested a more modest capacity of three (Broadbent, 1975) or four chunks (Coltheart, 1972; Cowan, 2001), more in line with this parameter.

After training, the 2AFC test is simulated by comparing the word and part-word values in the percept shaper. The stimulus with the higher value is the chosen one.

4.1.2.1 Parameters

PARSER has six parameters (Table 2), two of which are of outmost importance[7]: forgetting and interference. The most relevant feature of the model is the ratio between the increments (parameters 2, 3, and 6 of Table 2) and the decrements (parameters 4 and 5 of Table 2).

Parameter Number	Parameters	Standard Values
1.	Maximum number of shaping units used to form a percept	3
2.	Weight of initial shaping units	1
3.	Gain in weight for new percepts and percept reactivation	1
4.	*Forgetting*: loss in weight for every percept after each iteration	0.05
5.	*Interference*: loss in weight for the shaping units of the chosen percept of the current iteration	0.005
6.	Weight threshold above which a percept is able to shape perception	1

Table 7 - - Parameters of PARSER and respective standard values (as they appear in studies 1 and 2 of Perruchet & Vinter (1998) seminal paper and most of following studies featuring PARSER).

Forgetting

At each iteration the value of forgetting is subtracted to every percept in the percept shaper. It simulates natural decaying of stimuli memory traces.

Interference

If PARSER creates a new percept that, as shaping units, has another percept with more than one shaping unit, then the value of interference will be subtracted to the shaping units of this last percept. For example, if in the percept shaper there are the percepts KA, FUBI, and LA

[7] Perruchet et al (2014; page 4) call the other four "minor parameters" and in fact we are unaware of a study where they have been changed from the original values (e.g. see Table 5). We followed this trend and kept the same values in the present simulations.

and the new percept KAFUBILA is created, the shaping units FU and BI will be the object of interference. It simulates the difficulty of retrieving an item from memory when there are other items similar to this one in memory.

4.1.2.2 Simulation of older participants

PARSER has never been used to simulate older adults' performance. Since attention and working memory are impaired in older adults, it makes sense to increase the values of both Forgetting and Interference from the standard values: higher forgetting would simulate a lower working memory and higher interference higher cognitive noise. These increases will likely reflect a lower performance since these values will be subtracted from the percepts in the percept shaper.

4.2.3 Comparison in the literature

Although SRN and PARSER are the most representative models of each approach, their fit to the same human data has rarely been compared (Frank et al, 2010). It is thus important to compare these models with further behavioral data if we are to find the model that best simulates the cognitive mechanisms of statistical learning.

A recent study that compared both model with a paradigm not so different from the one used in the behavioral experiment of this work is that of Giroux & Rey (2009). Giroux & Rey (2009) exposed participants to an AL made of two trisyllabic and the four disyllabic words for either 2 or 10 minutes. After that they performed a two-alternative forced-choice with 16 trials: the four disyllabic were contrasted twice against different part-words and the trisyllabic words were divided into 4 sublexical disyllabic units, also contrasted twice against different part-words. They then feed the same AL to both models and assessed their performance.

		Giroux & Rey (2009)
SRN Parameters	l	0.2
	m	0.2
	a	14
	w	[0.5, 0.5]
	e	1
PARSER Parameters	1	3
	2	1
	3	1
	4	0.05
	5	0.005
	6	1
	e	1

Table 8 - Parameter values used in Giroux & Rey (2009). PARSER parameters are displayed by the same order as those in Table 2 (page X VER).

Parameter values can be seen in Table 3. For the two-alternative forced choice test the word with the biggest activation was chosen in the SRN and the word with the highest weight was chosen in PARSER (if only one was present that was chosen; if none, then the choice was random). Analyses of variance (ANOVA) were conducted using simulated participants as random variable, treating the type-of-item (words vs. sublexical units) as a within-participant variable and language duration (2 vs. 10 min) as a between-participant variable. PARSER seemed to be a better predictor of performance than the SRN (Figure 2).

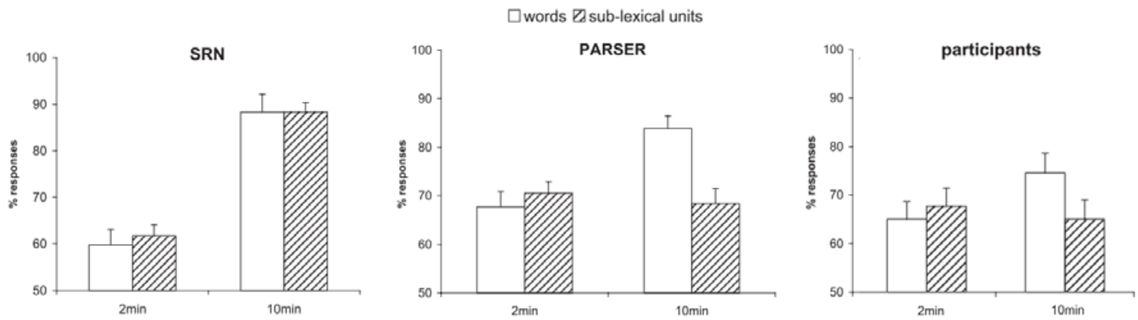


Figure 8 - Participants and models performance in the AL task of Giroux & Rey (2009).

4.3 Objectives and Hypotheses

The objective of this experiment was to compare the goodness of fitness of SRN and PARSER in order to understand which of the two approaches to speech segmentation, bracketing and clustering respectively, had more support from the model.

SRN is a somewhat complex model, with several parameters and based on neurological realist mechanisms. It is, perhaps, too efficient at extracting regularities from the environment, performing often well above human learners in speech SL modeling. Parser on the other hand is a very simple model with only two main parameter and based on the cognitive systems of attention and working memory. Despite its simplicity it has fit quite well some behavioral data of speech SL. The AL of Fernandes et al (2005) has, however, higher variability of TPs than other languages that have been used before in the comparison of these models. Still, it has very few words repeated for a very long period of time. This makes it difficult to make a prediction on which model will fit the best. It is even likely that a better model of human cognitive capacities for speech segmentation would fuse characteristics of these two.

5. Method

In order to simulate participants' performance both models were fed the whole language used on the behavioral study during a single iteration. 27 runs for each group (young, old) were performed in both models, simulating the participants from the behavioral experiment. The forced-choice test performed in the behavioral experiment was then simulated for both models.

5.1 Simple Recurrent Network

Simulations were run under Pybrain (Schaul et al, 2010), a python module for neural network construction, on Windows 8.1.

5.1.1 Model of young adults

We started by accessing the parameter values for optimal performance of the model. Human cognitive capacities are believed to peak in the first half of the 20s (Schaie 2005). The average age for the young participants in the behavioral experiment was almost 20 yo, around the age of optimal performance, so we took the liberty of setting the optimal parameters for the model of young adults[8]. Following previous studies, we tested the parameter space of the SRN model of young adults by systematically varying the number of hidden units (even numbers from 4 to 30), the learning range (.1; .2; .3) and the momentum (.1; .2; .3). Predictions of the model did not change qualitatively over the last two parameters, it was thus used the .1 value for both parameters. The model showed no increased of performance above 10 hidden units, so this value was chosen (Appendix 1)

We constructed a SRN with four layers: an input and an output layers with 11 nodes, one for each syllable in the language, and a hidden and context layers with 10 nodes (Fig. 3). Input nodes were fully connected to hidden nodes and hidden nodes fully connected to output nodes. Each hidden node was connected to one context node, to which copied its value. In the next iteration, that value was fed back to all hidden nodes. In this way, for each iteration hidden nodes receive input from the input layer (present syllable) and from the context layer (syllable before the present one). Random initial weights were assigned to each connection at the beginning of each run.

A supervised learning method was used where at time t_0 the 1st syllable would be used as input and the 2nd as target, at time t_1 the 2nd syllable would be used as input and the 3rd as target, and so on. The learning algorithm used was backpropagation through time (BPTT).

For the input layer we used a linear activation function, for the hidden layer a hyperbolic tangent activation function and for the output layer a softmax activation function.

[8] In connectionist models is usual to report only the simulation for the parameter set that more closely matched the human performance

In the test phase each syllable of the language was presented to the net and the probabilities of the next syllable were retrieved and compared with the correct probabilities (See Table 6 and Figure 4). Network performance for word and part-word was assessed as the activation of the output unit corresponding to the second syllable after the network had been presented with the first, and the third after the second was presented. By averaging both values we got the TP of each word and part-word. To simulate the two-alternative forced choice test the orthogonal combination of words and part-words were tested and the stimulus with the highest value was selected as the "winner". At the end of each run a .txt file with the networks parameters and the results of the two- alternative forced choice test was saved.

5.1.2 Model of older adults

Partially following Hinton & Shallice (1991), and having in mind the cognitive resources used for speech SL and how cognitive aging influences them, we decided to implement the following changes in the model of older adults:

(i) randomly setting some of the connections' weights to 0 after training;

(ii) decrease the value of learning rate, momentum and weight decay;

(iii) removing some of the neurons from the hidden layer;

(iv) skipping some of the input patterns.

We explored the parameter space by systematically varying the percentage of connections set to 0 and skipped patters (0%; 20%; 40%) and the learning rate and momentum (.06; .08; .1). Since we used 10 hidden nodes in the model of young adults and the average brain shrinkage from young adulthood to late adulthood is about 30%, we used 7 hidden nodes in the model of older adults. Everything else remained the same as the model of young adults. Neither the learning rate nor the momentum had a great impact on network performance (Appendix 2) so we decided to use the value of .08 for each. Skipped patterns and damaged connections seemed to be too severe at .4 for performance (Appendix 2), so we decided to use the values of .2 for each.

5.2 Parser

Simulations were run with U-learn (Perruchet, Robinet & Lemaire, submitted for publication; freely available at www.leadserv.u-bourgogne.fr/~perruchet) on Windows 8.1. We did 27 runs for young and older adults' models simulating the 27 participants of each age-group of the behavioral experiment. The seed was kept the same for both models. The values of the percepts in the percept shaper of each run were saved into a txt file. The 2AFC test was simulated by comparing the word and part-word values in the percept shaper.

We also evaluated whether PARSER would perform better for high- than low-TP-words as suggested by Perruchet & Vinter, (1998). Because the value of each chunk is absolute, to have a relative measure of comparison between runs and between low- and high-TP-words, for each run we summed the values of all percepts in the percept shaper for each run and set that value to 100% of the memory. We then summed the values of the three words of each TP-value group and calculated their percentage in relation to the total. After averaging for all the runs, these values were compared between models.

5.2.1 Model of young adults

We used the default parameters (see Table 2) for young participants' simulation, following the original model (Perruchet & Vinter, 1998) and subsequent modeling by the author (e.g. Perruchet et al 2014).

5.2.1 Model of older adults

We tested the parameter space of PARSER by systematically varying the values of decay and interference (.1; .05; .01; .005). For some combinations of these values there would be largely only monosyllabic percepts or polysyllabic percepts (3n, n>1; two or more times the size of a word) in the percept shaper, so these were discarded (Appendix X). The best performance was located around the values of .001 and .05 for decay and .01 for interference. We chose to average these values, reaching a rate of decay of .07 and a rate of interference of .01 for the model of older adults. All other parameters remained unchanged (see Table 5).

6. Results and Discussion

6.1 Simple Recurrent Network

There was no statistically significant difference between simulations of young and older for the same TP-group simulation (Figure 5): low-TP-word group, t(26) = -0.685, p = 0.497; high-TP-word group, t(26) = -2.318, p = 0.024. Difference between low- and high-TP-words were statistically significant for both age-groups: young group, t(26) = -14.006, p > 0.0001; old group, t(26) = -14.381, p > 0.0001. Results were very similar between age-groups, which might either indicate that the network variables manipulated were not sufficiently different or the calculations of the 2AFC were not representative of the ones done by the human participants.

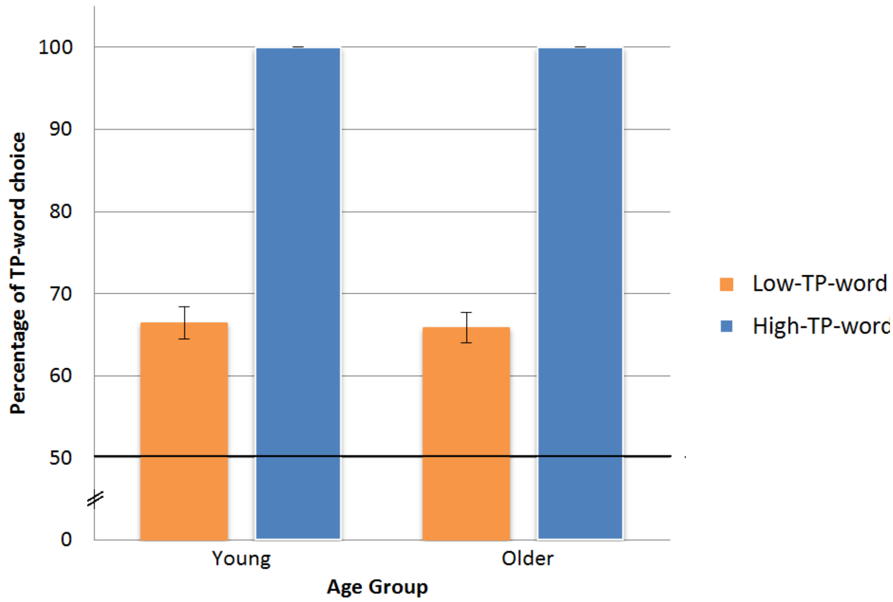


Figure 5 - Percentage of each TP-word group chosen by young and older participants in the 2AFC test SRN simulation.

6.2 PARSER

There was no sign of part-words in PARSER's percept shaper in either condition (e.g. Figure 6), which means that if we were to perform a 2AFC test as it is in the bibliography, PARSER would have a performance of 100% for both age groups, clearly above the one of human participants. Independently of the parameter values used, it would be near impossible to bring part-words into PARSERs memory since their frequency in the language is so small (Table 7). This indicates that PARSER mechanisms of speech segmentation are either not the correct ones or they play a smaller role than anticipated by their authors.

```
KA/FU/BI/.................. 83.92
FU/FI/BU/.................. 82.655
KI/LA/BU/.................. 73.465
BU/KA/LA/.................. 72.31
BA/BU/KU/.................. 59.125
LU/FA/BA/.................. 46.805
BU/........................ 19.86
BA/BU/KU/LU/FA/BA/.......... 7.375
KI/LA/BU/LU/FA/BA/BU/KA/LA/.. 1.03
......................................
FU/FI/BU/BU/KA/LA/.......... 0.94
BA/BU/KU/KI/LA/BU/.......... 0.875
KI/LA/BU/FU/FI/BU/.......... 0.78
KI/LA/BU/BU/KA/LA/FU/FI/BU/.. 0.525
FU/FI/BU/BA/BU/KU/LU/FA/BA/.. 0.465
BU/KA/LA/FU/FI/BU/.......... 0.295
KA/FU/BI/KI/LA/BU/KA/FU/BI/.. 0
```

```
FU/FI/BU/.................................. 96.574
KI/LA/BU/.................................. 77.604
KA/FU/BI/.................................. 76.586
BU/KA/LA/.................................. 75.012
BA/BU/KU/.................................. 57.566
LU/FA/BA/.................................. 35.539
BA/BU/KU/LU/FA/BA/........................... 15.01
KA/FU/BI/KI/LA/BU/........................... 7.53
BU/KA/LA/LU/FA/BA/........................... 5.686
BA/........................................ 2.926
..................................................
BA/BU/KU/KI/LA/BU/FU/FI/BU/................... 0.935
KA/FU/BI/FU/FI/BU/........................... 0.83
BU/KA/LA/KI/LA/BU/FU/FI/BU/................... 0.786
FU/FI/BU/KI/LA/BU/........................... 0.705
BA/BU/KU/LU/FA/BA/KA/FU/BI/KI/LA/BU/BU/KA/LA/.. 0.444
LU/FA/BA/BU/KA/LA/LU/FA/BA/................... 0.424
FU/FI/BU/BU/KA/LA/........................... 0.405
FU/FI/BU/KI/LA/BU/FU/FI/BU/................... 0.379
KI/LA/BU/KA/FU/BI/........................... 0.245
```

Figure 6 - Example of each percepts and the corresponding value in PARSER's percept shaper after a run simulating a young (left) and older (right) participants. Note that there is no part-words in the percept shape for either simulation.

Part-Word	Frequency of occurrence in the AL	Percentage of occurrence in the AL
la-bu#fu	51	0,024
ku#bu-ka	24	0,011
ka-la#fu	48	0,022
fi-bu#lu	3	0,001
ba#ki-la	9	0,004
bi#lu-fa	12	0,006

Table 1 - Frequency and percentages of occurrences of the part-words in the language

Then we compared the content of the percept shaper to see if PARSER could indeed present higher values in memory for high-TP-words compared to the low-TP-words, as defendend by the authors (Perruchet et al., 2014). Although high-TP-words values were higher than the ones of low-TP-words, no statistically significant difference between low- and high-TP-words was found in either age-group simulation (Fig. 7): young group, t(26) = -.914, p = .185; old group, t(26) = -.688, p = .737. The difference between young and older for the same TP-group simulation was significant in both cases: low-TP-word group, t(26) = 4,17, p = > .001; high-TP-word group, t(26) = -3,169, p = > .001.

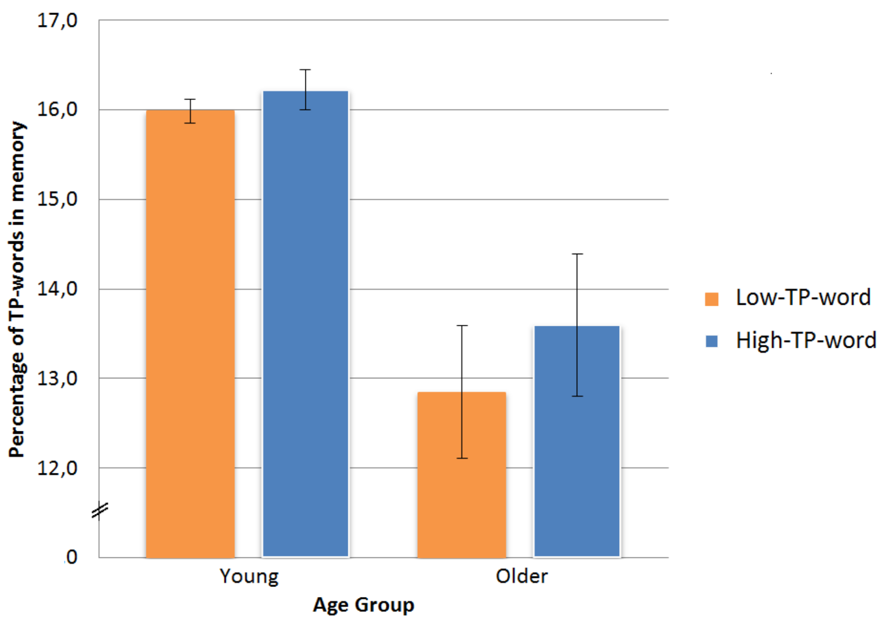


Figure 7 - Percentage of each TP-word group in PARSER's memory for each condition.

Because PARSER takes mainly in account chunk frequency, the same for low- and high-TP-words, it is understandable that there is no difference in performance between them.

6.3 Considerations for both models

It is well known that prior experience influences present learning. Both models started the simulation as a "blanck slate", which is obviously a very different case from the human participants. Lack of prior experience might be the cause to the much greater performance of the models compared with human participants.

7. Conclusion

The present work provided some advances in our understanding of the cues implicated in statistical speech segmentation in older adults. Syllables transitional probabilities seem to preserve its importance as a cue across life, providing that they are salient enough to overcome the decline of some cognitive systems. The brain areas responsible for statistical learning suffer a decreasing and their connections change across life. Furthermore, the capability for attentional sustenance and allocation, as well as working memory, gets increasingly compromised with aging. The lower performance of older participants was then expected, but a behavioral experiment alone is not enough to pin down the exact causes of this decline. Future neuroimaging studies one would help to better understand the interaction between these variables.

Studies of SL in older adults can have, besides the scientific value, a practical aspect for healthy aging. SL is related to language performance (Daltrozzo et al., 2013), so by understanding the mechanisms of SL in older adults it would be possible to implemented training programs to promoting active aging.

Some aspects of the methodology of the behavioral study could be improved. Some studies of speech SL set a postexperimental interview where the participants are asked what they were doing while listening or whether they noticed any repeating patterns during the familiarization phase and which would they be (e.g. Schapiro et al., 2014). This is useful to further disentangle the contribution of implicit and explicit systems to SL and could have been implemented in our study as well. Another aspect has to do with the rate of presentation of the AL. It has been pointed that older adults' speak about 20% slower than younger adults (Amerman & Parnell, 1992) and have more difficulty understanding rapidly presented speech

(Wingfield et al., 1999). To be sure that the results of the behavioral study were not caused by this perceptual difficulties, it would be advisable to redo the experiment with a (around 20%) slower rate of presentation of the AL.

This work also helped to access the biological validity of SRN and PARSER in regards to a speech SL task. Both models outperformed human performance, but SRN by a lesser degree, indicating, even if in a very feeble way, that the mechanisms behind it might be the same in operation for speech SL. Even if a couple of decades have passed since SRN and Parser have been designed, they have been subjected to very few changes until today in the context of computational modeling studies of speech SL. The increase in processing power and origin and growth of new simulation environments provide the opportunity to update these models, while overperformance, registered here and elsewhere, should provide the motive.

8. Bibliography

Amerman, J. D., & Parnell, M. M. (1992). Speech timing strategies in elderly adults. *Journal of Phonetics*, 20, 65–76.

Anderson, M. C. (2003). Rethinking interference theory: executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49, 415–445.

Anderson, J.R. (1983). The Architecture of Cognition, Harvard University Press.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.

Baddeley, A. D. (1986). Working Memory. Oxford, UK: Oxford Univ. Press.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, 15, 460–466.

Bapi, R. S., Chandrasekhar Pammi,V. S.,Miyapuram, K. P., & Ahmed, A. (2005). Investigation of sequence processing: a cognitive and computational neuroscience perspective. *Current Science*, 89, 1690–1698.

Bates, E. & Elman, J. (1996). Learning rediscovered. *Science*, 274, 1849–1850.

Bishop, N. A., Lu, T., & Yankner, B. A. (2010). Neural mechanisms of ageing and cognitive decline. *Nature*, 464, 529–535.

Bo, J., & Seidler, S. (2012). Differential working memory correlates for implicit sequence performance in young and older adults, 467–477.

Borella, E., Carretti, B., & De Beni, R. (2008). Working memory and inhibition across the adult life-span. *Acta Psychologica*, 128, 33–44.

Braver, T. S., & West, R. L. (2008). Working memory, executive processes, and aging. In Handbook of Aging and Cognition, 3rd Ed., eds F. I. Craik, and T.L. Salthouse (New York, NY: Lawrence Erlbaum Associates), 311–372.

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, 21, 399–405.

Castro, S. L., Cunha, L. S., & Martins, L. (2000). Teste Stroop Neuropsicológico em Português. Disponibilizado por Laboratório de Fala da Faculdade de Psicologia da Universidade do Porto em http://www.fpce.up.pt/labfala [2014, 09, 01].

Chomsky, N. (1980). Rules and representations. New York: Columbia University Press.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to Segment Speech Using Multiple Cues: A Connectionist Model. *Language and Cognitive Processes*, 13, 221–268.

Clapp, W. C., & Gazzaley, A. (2010). Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiology of Aging*, 33, 134–148.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31, 24–39.

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sci- ences*, 1145, 113–131.

Courchesne, E., Chisum, H.J., Townsend, J., Cowles, A., Covington, J. & Egaas, B. (2000). Normal brain development and aging: Quantitative analysis at in vivo MRI imaging in healthy volunteers. *Radiology*, 216, 672–682.

Craik, F. I. M., & Salthouse, T. A. (Eds.). (2008). The handbook of aging and cognition (3rd ed.). Mahwah, NJ: Erlbaum

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*, 48, 541–53.

Daltrozzo J., Conway C. M., & Smith G. N. L. (2013). Rehabilitating language disorders by improving sequential processing: a review. *The Journal of Macro Trend in Health and Medicine*, 1, 41-57.

Dennis, N. A., & Cabeza, R. (2008). Neuroimaging of healthy cognitive aging. In F. I. M. Craik & T. A. Salthouse (Eds.), Handbook of aging and cognition (3rd edition). Mahwah, NJ: Erlbaum.

Durrant, S. J., Cairney, S. A, & Lewis, P. A. (2012). Overnight consolidation aids the transfer of statistical knowledge from the medial temporal lobe to the striatum. *Cerebral Cortex*, 23, 2467–78.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & van der Vreken, O. (1996). The MBROLA Project : Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes. Proceedings of ICSLP'96, Philadelphia, 3, 1393–1396.

Eichenbaum, H. (2000). A cortical–hippocampal system for declarative memory. *Nature Reviews Neuroscience*, 1, 41–50.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.

Elman, J. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48, 71-99.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, 64, 1021–1040

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105, 247–299.

Fernandes, T. (2007). From the acoustic stream to the spoken word: Stages of processing of speech perception. [Da corrente acoustica à palavra falada: Estádios do processamento da percepção da fala]. University of Lisbon, Thesis.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics*, 72, 1522–1532.

Fernandes, T., Ventura, P., & Kolinsky, R. (2007). Statistical information and coarticulation as cues to word boundaries: a matter of signal quality. *Perception & Psychophysics*, 69, 856–64.

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. Psychological Science, 12(6), 499–504.

Fodor, J. A. (1975). The language of thought. Harvard University Press: Cambridge, MA.

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, 117, 107–25.

Freeman, W.J., Ed. (1975), Mass Action in the Nervous System. New York, Academic Press.

Freitas, S., Simões, M. R., Santana, I., Martins, C., & Nasreddine, Z. (2013). Montreal Cognitive Assessment (MoCA): Versão 1. Coimbra: Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: a recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, 118, 614–636.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, 22, 95–110.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360, 815–36.

Gathercole, S.E., Willis, C.S., Baddeley, A.D., & Emslie, H. (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory,* 2, 103-127.

Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, 8, 1298–1300

Gazzaley, A., & D'Esposito, M. (2007). Top-down modulation and normal aging. *Annals of the New York Academy of Sciences*, 1097, 67–83.

Gibson, E.J. (1991). An odyssey in learning and perception. Cambridge, MA: MIT Press.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, 33, 260–272.

Goodsitt, J.V., Morgan, J.L., & Kuhl, P.K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language,* 20, 229–252.

Goschke, T. (1998). Implicit learning of perceptual and motor sequences: Evidence for independent learning systems. In M. A. Stadler & P. A. Frensch (Eds.), Handbook of implicit learning (pp. 410–444). London: Sage.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B*, 290, 181–197.

Gruart, A., Muñoz, M. D., & Delgado- García, J. M. (2006). Involvement of the CA3-CA1 synapse in the acquisi- tion of associative learning in behavingmice. *Journal of Neurosciences*, 26, 1077–1087.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and new view. In G. H. Bower (Ed.), The psychology of learning and motivation, 22, 193–225. San Diego, CA: Academic Press.

Hebb, D.O. (1949). The Organization of Behavior. New York: Wiley & Sons.

Heimbauer, L. A., Conway, C. M., Christiansen, M. H., Beran, M. J., & Owren, M. J. (2010). Grammar rule-based sequence learning by rhesus macaques (Macaca mulatta). *American Journal of Primatology*, 72, 65

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74–95.

Howard D. V., & Howard J. H., Jr. (1989). Age differences in learning serial patterns: direct versus indirect measures. *Psychological Aging*, 4, 357–364.

Howard D. V., & Howard J. H., Jr. (1992). Adult age differences in the rate of learning serial patterns: evidence from direct and indirect tests. *Psychological Aging*, 7, 232–241.

Howard, D. V, Howard, J. H., Dennis, N. a, LaVine, S., & Valentino, K. (2008). Aging and implicit learning of an invariant association. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 63, 100–5.

Jääskeläinen, I. P., Ahveninen, J., Andermann, M. L., Belliveau, J. W., Raij, T., & Sams, M. (2011). Short-term plasticity as a neural mechanism supporting memory and attentional functions. *Brain Research*, 1422, 66–81.

Jacobs, B., Driscoll, L., & Schall, M. (1997). Life-span dendritic and spine changes in areas 10 and 18 of human cortex: a quantitative Golgi study. *Journal of Comparative Neurology*, 386, 661–680.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology*, 54, 1105–1124.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology*, 126, 278–287.

Karuza, E. a, Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, 127, 46–54.

Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, 110, 316–339.

Kemper, T. L. (1994). Neuroanatomical and neuropathological changes during aging and de-mentia. In M. L. Albert & J. E. Knoefel (Eds.), Clinical neurology of aging, 2, 3–67. New York: Oxford University Press.

Kersten, A., & Earles, J. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, 44, 250–273.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83, 35–42.

Kril, J., Hodges, J., & Halliday, G. (2004). Relationship between hippocampal volume and CA1 neuron loss in brains of humans with and without Alzheimer's disease. *Neuroscience Letters*, 361, 9-12.

Krogh, L., Vlach, H. A., & Johnson, S. P. (2012). Statistical learning across development: flexible yet constrained. *Frontiers in Psychology*, 3, 598.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of nonlocal musical rules: implicitly learning more than chunks. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31, 1417–32.

Lezak, M. D., Howieson, D. B., Bigler, E. D., and Tranel, D. (2012). Neuropsychological Assessment, 5th Edn.New York, NY: Oxford University Press.

Li, S.-C., and Lindenberger,U. (1999). Cross-level unification: A computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In L.-G.Nilsson and H.Markowitsch (eds), Cognitive Neuroscience ofMemory (pp. 104–146). Toronto:Hogrefe and Huber.

Lindenberger, U., & Oertzen, T. von (2006). Variability in cognitive cognitive aging: From taxonomy to theory. In F. I. M. Craik & E. Bialystok (Eds.), Lifespan cognition: Mechanisms of change (pp. 297–314). Oxford, UK: Oxford University Press.

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypothesis. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 812–817.

Marsh, J., Hughes, R., & Jones, D. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, 58, 682-700.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology*, 134, 477–500.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, 26, 7629–39.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 23, 1007–1028.

Mielke, J. (2008). The emergence of distinctive features. Oxford: OUP.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review*, 63, 81-97.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190–204.

Negash, S., Howard, D. V., Japikse, K. C., & Howard, J. H. Jr. (2003). Age-related differences in implicit learning of non-spatial sequential patterns. *Aging, Neuropsycholpgy and Cognition*, 10, 108–121.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science*, 14, 11-28.

O'Sullivan, M., Jones, D., Summers, P., Morris, R., Williams, S., & Markus, H. (2001). Evidence for cortical disconnection as a mechanism of age-related cognitive decline. *Neurology*, 57, 632–638

Onnis, L., Monaghan, P., Richmond, K., & Chater, K. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53, 225-237.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annual Review of Psychology*, 60, 173–96.

Pashler, H. E. (1998). The Psychology of Attention. Cambridge,MA: MIT Press

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–8.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics*, 17, 97–119.

Perruchet, P., & Tillmann, B. (2010). Exploiting multiple sources of information in learning an artificial language: human data and modeling. *Cognitive Science*, 34, 255–85.

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language*, 39, 246–263.

Perruchet, P., Poulin-Charronnat, B., Tillmann, B., & Peereman, R. (2014). New evidence for chunk-based models in word segmentation. *Acta Psychologica*, 149, 1–8.

Perruchet, P., Robinet, V., & Lemaire, B. (2014). U-Learn: Finding optimal coding units from unsegmented sequential databases. (submitted for publication).

Poldrack, R. A, & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia*, 41, 245–251.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top- down influences. *Journal of Neuroscience*, 26, 4970–4982.

Rabbitt, P., Lunn, M., Pendleton, N., Horan, M., Scott, M., Thacker, N., Lowe, C., & Jackson, A. (2007a). White matter lesions account for all age-related declines in speed but not in intelligence. *Neuropsychology*, 21, 363–370 .

Rabbitt, P., Mogapi, O. , Scott, M., Thacker, N., Lowe, C., Horan, M., Pendleton, N., Jackson, A., & Lunn, D. (2007b). Effects of global atrophy, white matter lesions, and cerebral blood flow on age-related changes in speed, memory, intelligence, vocabulary, and frontal function. *Neuropsychology*, 21, 684–695 .

Raz, N., Lindenbeeger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., & Williamson, A. (2005). Regional brain changes in aging healthy adults: general trends, individual differences, and modifiers. *Cereberal Cortex*, 15, 1676–1689.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behaviour*, 6, 855–863.

Rieckmann, A., & Bäckman, L. (2009). Implicit learning in aging: extant patterns and new directions. *Neuropsychology Review*, 19, 490–503.

Robinet, V., Lemaire, B., & Gordon, M. B. (2011). MDLChunker: a MDL-based cognitive model of inductive learning. *Cognitive science*, 35, 1352–89.

Rohde, D.L.T. & Plaut, D.C. (2002). Less is less in language acquisition. In Quinlin, P.(Ed.),Connectionist modeling of cognitive development. Hove, UK: Psychology Press.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Cognitive Science*, 1, 906–914.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by back-propagating errors. *Nature*, 323, 533– 536.

Saffran, J.R., Aslin, R.N., & Newport, E.L., (1996a). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.

Saffran, J.R., Newport, E.L., & Aslin, R.N., (1996b). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35, 606–621.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.

Salthouse, T. A. (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society, 16, 754–60.

Schaie, K.W. (2005). Developmental influences on adult intelligence: The Seattle Longitudinal Study. New York: Oxford University Press.

Schapiro A. C., Gregory E., Landau B., McCloskey M., & Turk-Browne N. B. (2014). The necessity of the medial temporal lobe for statistical learning. Journal of Cognitive Neuroscience, 26, 1736–1747.

Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Ruckstieß, & J. Schmidhuber (2010). Pybrain. *Journal of Machine Learning Research*, 11, 743–746.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime references guide. Pittsburgh: Psychology Software Tools Inc.

Seger, C. A. (1997). Two forms of sequential implicit learning. Consciousness and Cognition, 6, 108–131.

Servan-Schreiber, E., & Anderson, J. (1990). Learning artificial grammars with competitive chunking. Journal of Experimental Psychology. *Learning, Memory, and Cognition*, 16, 592–608.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual Sequence Learning in Infancy: Domain-General and Domain-Specific Associations with Language. Infancy, 17, 247–271.

Siegelmann, H. T. (1999). Neural networks and analog computation: Beyond the turing limit. Basel: Birkhäuser.

Slifkin, A. B., & Newell,K.M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science*, 7, 170–177.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643.

Stuss, D. T., Alexander, M. P., Palumbo, C. L., Buckle, L., Sayer, L., & Pogue, J. (1994). Organizational strategies of patients with unilateral or bilateral frontal lobe injury in word list learning tasks. Neuropsychology, 8, 355-373.

Sullivan, E. V., Rosenbloom, M., Serventi, K. L., & Pfefferbaum, A. (2004). Effects of age and sex on volumes of the thalamus, pons, and cortex. *Neurobiology of Aging*, 25, 185–192.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, 50, 86–132.

Szalma, J. L., & Hancock, P. A. (2011). Noise effects on human performance: a meta-analytic synthesis. American Psychological Association, 137, 682-707.

Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: a mathematical model of distributional statistical learning. *Cognitive Science*, 37, 310–43.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.

Thiessen, E. D., (2009). Statistical learning. In Bavin, E. (ed.), The Cambridge handbook of child language. Cambridge: Cambridge University Press, 2009.

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, 67, 867–875.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, 21, 1934–45.

Turken, A. U., Whitfield-Gabrieli, S., Bammer, R., Baldo, J. V., Dronkers, N. F., & Gabrieli, J. D. E. (2008). Cognitive processing speed and the structure of white matter pathways: Convergent evidence from normal variation and lesion studies. *Neuroimage,* 42, 1032-1044.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging*, 13, 120–126.

Wechsler, D. (1997). WAIS-III administration and scoring manual. San Antonio, TX: The Psychological Corporation.

West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, 37, 179– 189.

Wingfield, A., & Ducharme, J. L. (1999) Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *Journal of Gerontology*, 54, 199–202.

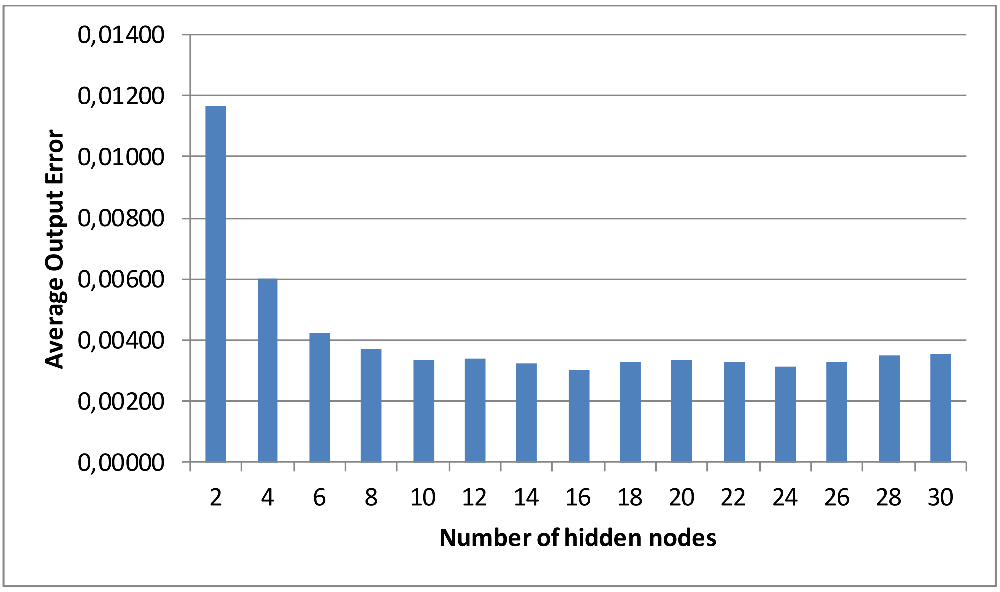
Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, 8, 451–456.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: integrating statistical and social cues. *Neurocomputing*, 70, 2149–2165.

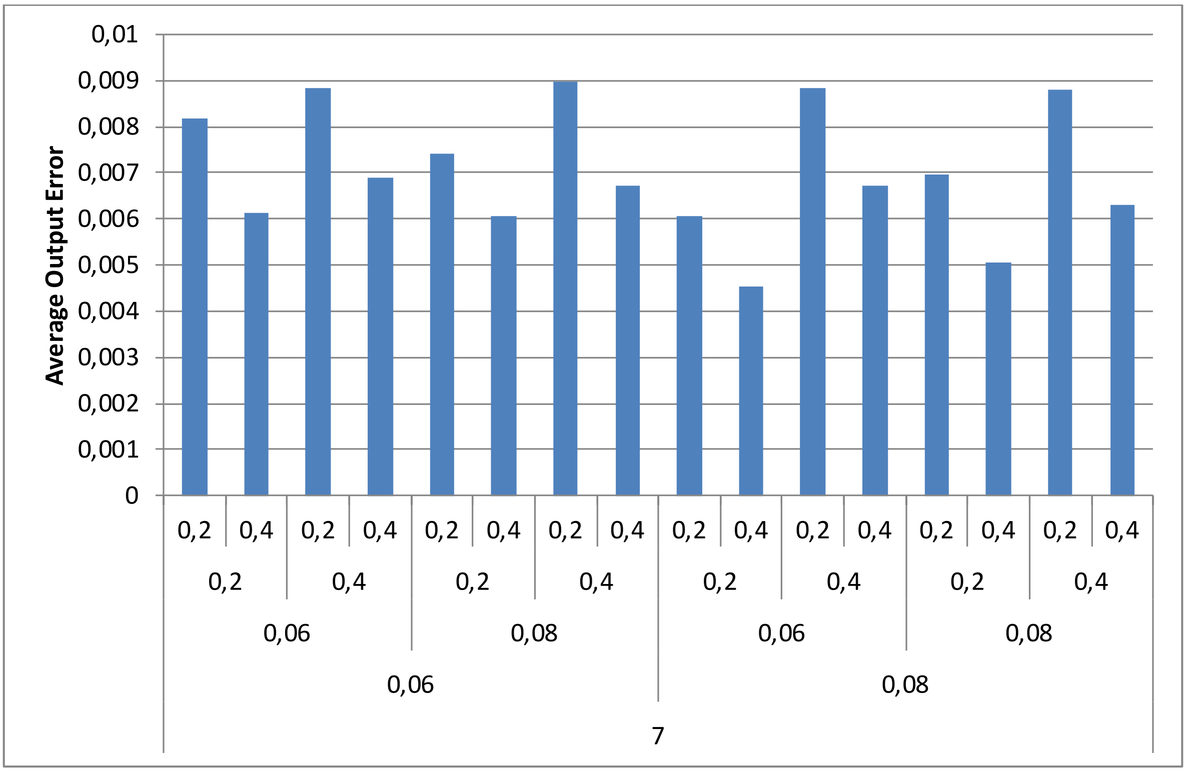
Zanto,T.P., & Gazzaley, A. (2014). Attention and Aging. In Nobre, A.C. and Kastner, S. (Eds), The Oxford Handbook of Attention.(pp. 927-971). Oxford, UK: Oxford University Press. (2014)

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, 24, 667–77.

Appendix



Appendix 1 - Graph showing the relation between the number of hidden nodes and the average output error of the young adult's simulation



Appendix 2 - Graph showing the relation between five variables and the output error. The variables are, from bottom to top: number of hidden nodes, learning rate, momentum, percentage of attention deficit and percentage of damaged conections.